REPORT DOCUMENTATION PAGE		Form Approved OMB NO. 0704-0188

Public Reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comment regarding this burden estimates or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188,) Washington, DC 20503.

1. AGENCY USE ONLY (Leave Blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED Final Report 7/1/2002-6/30/2004
4. TITLE AND SUBTITLE Tunable Snapshot Spectrometer Feasibility Study		5. FUNDING NUMBERS DAAD19-02-1-0265
6. AUTHOR(S) Christopher P. Tebow Eustace L. Dereniak		
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) University of Arizona 1630 E. University Blvd. Tucson, AZ 85721		8. PERFORMING ORGANIZATION REPORT NUMBER IR1
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) U. S. Army Research Office P.O. Box 12211 Research Triangle Park, NC 27709-2211		10. SPONSORING / MONITORING AGENCY REPORT NUMBER 44042.19-PH

11. SUPPLEMENTARY NOTES
    The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision, unless so designated by other documentation.

12 a. DISTRIBUTION / AVAILABILITY STATEMENT Approved for public release; distribution unlimited.	12 b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)

A tunable snapshot imaging spectrometer has been demonstrated. A liquid crystal spatial light modulator (LC SLM) has been integrated into a computed tomographic imaging spectrometer (CTIS) to achieve tunability. The LC SLM allows for rapid, programmable, and non-mechanical alteration of its phase profile by the application of appropriate voltages to its transparent electrodes.

The goal of this dissertation is twofold: (1) to integrate a liquid crystal spatial light modulator into a CTIS instrument and characterize its performance as a tunable CTIS disperser, and (2) to implement tunability by analyzing different CTIS configurations.

The theoretical model of CTIS operation, calibration, reconstruction, and disperser design are covered in detail. The cross talk of the LC SLM forces the use of a feedback design algorithm rather than designing the desired phase profile a priori in the computer. The modifications to the current polychromatic linear inversion technique for use with the LC SLM in feedback are presented. The result of the modifications is the successful integration of a reprogrammable (i.e. tunable) disperser for the CTIS instrument.

The implementation of tunability is explored by analyzing the spectral resolution of a reconstructed point source for different disperser configurations. A method for experimentally determining the CTIS spectral resolution is presented.

14. SUBJECT TERMS Computed Tomographic Imaging Spectrometer Liquid Crystal Spatial Light Modulator Computer Generated Hologram			15. NUMBER OF PAGES 138
			16. PRICE CODE
17. SECURITY CLASSIFICATION OR REPORT **UNCLASSIFIED**	18. SECURITY CLASSIFICATION ON THIS PAGE **UNCLASSIFIED**	19. SECURITY CLASSIFICATION OF ABSTRACT **UNCLASSIFIED**	20. LIMITATION OF ABSTRACT **UL**

TUNABLE SNAPSHOT IMAGING SPECTROMETER
ARO No. DAAD 19-02-1-0265
UofA FRS # 327490

by
Christopher Paul Tebow

THE UNIVERSITY OF ARIZONA

Due September 30, 2004

TABLE OF CONTENTS

TABLE OF CONTENTS - CONTINUED

LIST OF FIGURES

LIST OF FIGURES – CONTINUED

LIST OF FIGURES – CONTINUED

LIST OF FIGURES – CONTINUED

LIST OF FIGURES – CONTINUED

LIST OF FIGURES – CONTINUED

LIST OF FIGURES – CONTINUED

ABSTRACT

A tunable snapshot imaging spectrometer has been demonstrated. A liquid crystal spatial light modulator (LC SLM) has been integrated into a computed tomographic imaging spectrometer (CTIS) to achieve tunability. The LC SLM allows for rapid, programmable, and non-mechanical alteration of its phase profile by the application of appropriate voltages to its transparent electrodes.

The goal of this dissertation is twofold: (1) to integrate a liquid crystal spatial light modulator into a CTIS instrument and characterize its performance as a tunable CTIS disperser, and (2) to implement tunability by analyzing different CTIS configurations.

The theoretical model of CTIS operation, calibration, reconstruction, and disperser design are covered in detail. The cross talk of the LC SLM forces the use of a feedback design algorithm rather than designing the desired phase profile a priori in the computer. The modifications to the current polychromatic linear inversion technique for use with the LC SLM in feedback are presented. The result of the modifications is the successful integration of a reprogrammable (i.e. tunable) disperser for the CTIS instrument.

The implementation of tunability is explored by analyzing the spectral resolution of a reconstructed point source for different disperser configurations. A method for experimentally determining the CTIS spectral resolution is presented.

CHAPTER 1 - IMAGING SPECTROMETRY

The goal of imaging spectrometry is the acquisition of a three-dimensional data set, called the *data cube*. The dimensions of the data cube are the two spatial dimensions of an image (x,y) and a third wavelength (λ) dimension. The data cube is denoted as $\hat{f}(x, y, \lambda)$, where the 'hat' notation is used to signify that the data cube is an estimate of a physically real quantity called the *object cube*, which is denoted as $f(x, y, \lambda)$. The value of the data cube at any (x_o,y_o,λ_o) point is the relative spectral radiance of the corresponding point in the object cube. However, any real instrument has finite spatial and spectral resolutions, so the value of the data cube at (x_o,y_o,λ_o) is actually a relative spectral radiance *average* of a volume element, or voxel, of the continuous object cube whose volume is $V = \Delta x * \Delta y * \Delta\lambda$.

(a)   (b)   (c)
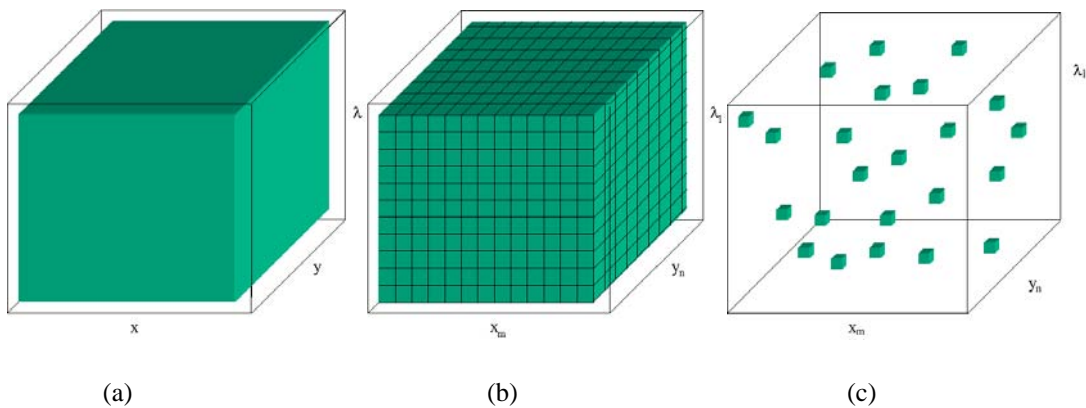
Figure 1.1: (a) The object cube, (b) discretized data cube, and (c) individual voxels of the data cube.

Digital imaging is inherently continuous-to-discrete, so the data cube is a discrete set of values comprising an estimate of the continuous, real object cube. Thus the data cube is better denoted as $\hat{f}(x_m, y_n, \lambda_l)$, where there is M x N x L samples of the object cube and the subscripts denote the particular voxel in question.

The most common example of a data cube is a color digital image. Each pixel of the image contains a red, green, and blue component comprising the composite color image. The focal plane arrays (FPAs) of commercial digital cameras detect the three different colors by using filters such that each pixel integrates over a specific wavelength band.
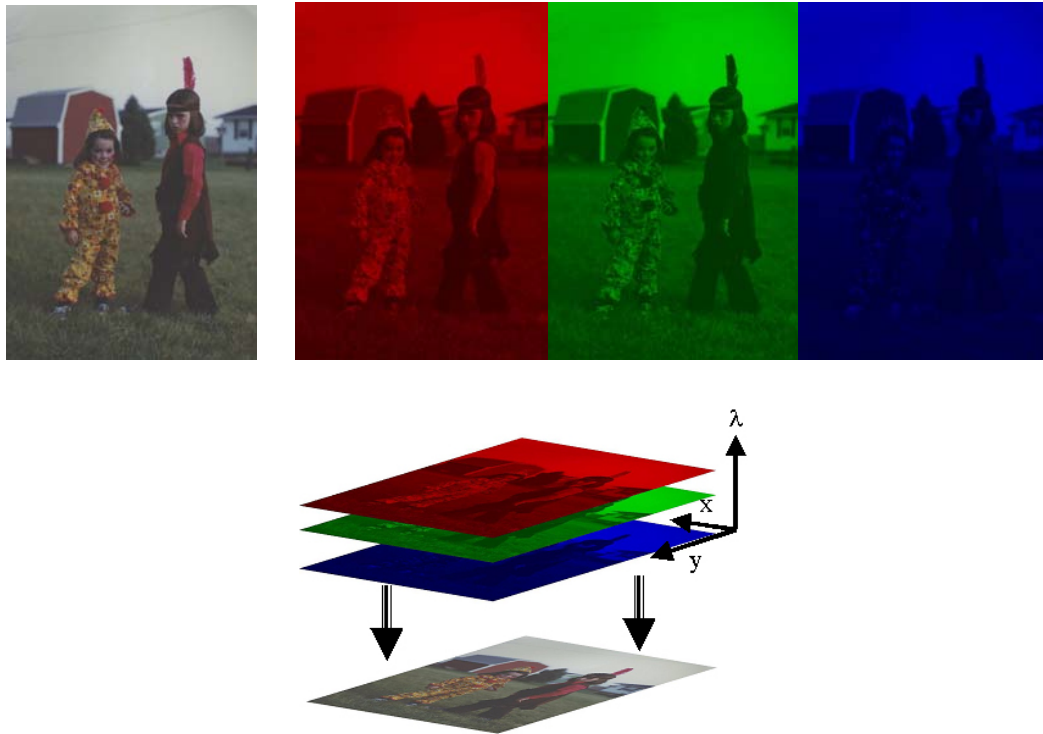
Figure 1.2: Data cube representation of the RGB components of a digital image.

The RGB image is a stack of three-color images that are easily separated on a home computer for viewing. This is the most common way of visualizing the data cube, as a stack of single color images. Three bands are quite sufficient for human visualization using a tri-phosphor CRT monitor, LCD display, or color printer, each of which display (or reflect) almost all of the colors on the CIE color chart within the range of human vision. However, a three-band data cube is insufficient for scientific studies and feature identification. Multi-spectral imaging spectrometers collect a stack of several bands, while the number of bands for hyperspectral instruments is in the hundreds. The fundamental limitations on the number of bands that an instrument can collect are the instrument dispersion, throughput, and FPA pixel size and sensitivity.

Another way to view the data cube is to regard it as a set of hundreds or thousands of individual point spectra collected simultaneously. The data cube is then a set of columns (an example of which is the whiskbroom data of Figure 1.4a.) corresponding to each spatial (x_n, y_m) pixel rather than a stack of x-y planes. The combination of imaging and spectrometry is a powerful remote sensing tool, and the choice of how best to view the data cube is based on application and convenience.

Acquiring the data cube

Almost all imaging spectrometers require some type of scanning to acquire the data cube, and most require moving parts. This is because current FPA technology is limited to a two-dimensional format of pixels that attempts to acquire a three-dimensional object. Without scanning or beam division, the spatial and spectral information of the

data cube overlap and the content from several pixels are spread into each particular voxel: i.e., there is spatial/spectral cross talk. One can imagine a three dimensional FPA which consists of a stack of planes of pixels, each of which absorbs only a very small band of wavelengths and passes all others. Such an FPA would preclude the need for scanning and simplify the basic imaging spectrometer. A 'focal plane cube' such as this is still in the realm of imagination; however, FOVEON has developed a 3-band visible array that collects the RGB components of the scene in each pixel.[1] There is also a significant amount of work currently in progress to fabricate dual band cameras where one array absorbs 3-5 um light and the other absorbs 8-12 um light.

Some of the more common imaging spectrometers include whiskbroom and push broom scanners, filtered cameras, wedge spectrometers, Fourier transform spectrometers, and computed tomographic imaging spectrometers.

An example of a whiskbroom scanner is the Airborne Visible and Infra Red Imaging Spectrometer[2] (AVIRIS) for earth remote sensing in the 400 – 2500 nm spectral range using 224 bands. The AVIRIS uses a diffraction grating, a linear array of 224 pixels, and a scanning mirror to sweep the field of view to obtain a 614 pixel line of data with 10 nm bandwidth. The instrument is flown on a NASA ER-2 (an updated version of the famous U2 spy plane) and the along track motion provides the scanning along the other spatial dimension. Push broom scanners work in much the same way, only rather than a linear array they use a 2D array and do not need to scan in the cross-track dimension.
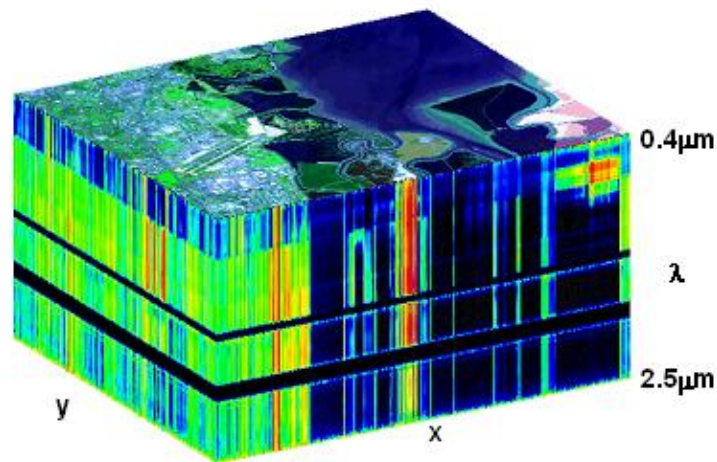
Figure 1.3: The AVIRIS data cube of Moffet field.[2]

Filtered cameras scan along the spectral dimension of the data cube. A rotating spectral filter wheel is placed in the optical train of a 2D staring FPA that collects full spatial resolution images as the wheel spins.

Wedge spectrometers[3] use a 'wedge' thin film that acts as a narrow band filter with a varying center wavelength along the displacement axis of the wedge. The filter can be fabricated directly onto a 2D FPA and the images taken form a tilted plane in the data cube. The scene is subsequently scanned in the direction of the wedge to acquire the entire data cube.

Fourier transform spectrometers[4] are based on the Michelson interferometer and encode the spectral information of a point to an interferogram. The instrument has very good signal-to-noise ratio over a broad range of spectra, but is also very sensitive to environmental conditions because it is an interferometer. The field of view of these

instruments is limited by the ability to keep a consistent optical path difference for off axis points.

Computed tomographic imaging spectrometers are a class of instruments that allow the spatial and spectral information of the data cube to overlap. In the Computed Tomographic Hyperspectral Imaging Spectrometer[5], or CTHIS, a direct view prism is used to disperse an image of a scene. Multiple images are taken while rotating the prism. The resulting data is a set of 2D projections of the 3D data cube in the same way a CAT scan forms a set of 2D images of a spatial, 3D object. A post-processing step using complex algorithms is needed to reconstruct the 3D data cube from its set of projections.

The biggest challenge with scanning imaging spectrometers is spatial and spectral misregistration. Often either the platform, as in the case of the AVIRIS, or the object is moving so that by the time one column or plane of the data cube is acquired the scene has changed when acquisition for the next column or plane has started. This problem can be averted with a snapshot imaging spectrometer where the data for the entire data cube is collected in one integration time of the FPA.
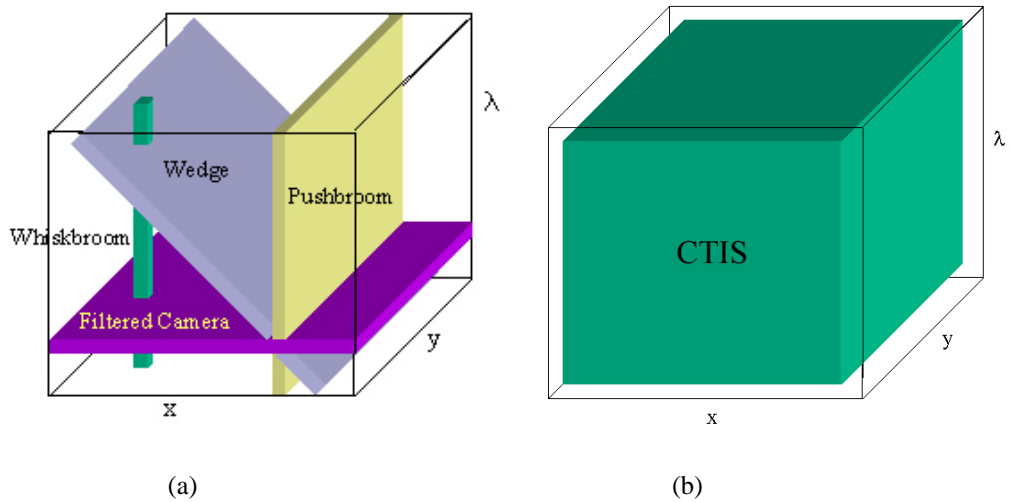
Figure 1.4: (a) Regions of the data cube collected in a single snapshot with some of the most common imaging spectrometers compared with (b) collecting the entire data cube in a single integration time with the CTIS.

Snapshot acquisition

The Computed Tomographic Imaging Spectrometer[6] (CTIS) is a snapshot imaging spectrometer that does not use any scanning and has no moving parts. As with the CTHIS system mentioned previously, the spatial and spectral information of the scene is allowed to overlap in a set of 2D projections of the data cube. The CTIS, however, collects the entire set of projections in a single frame. The CTIS uses a diffraction grating to diffract the central field of view into multiple orders, each of which is dispersed in wavelength. The field of view of the CTIS is minified, leaving room on the rest of the FPA for collection of these higher diffraction orders. Figure 1.5 shows an example of the raw data collected with the CTIS from a broadband, extended source.
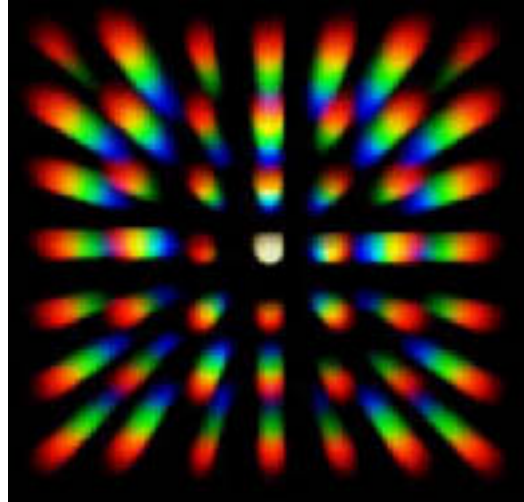
Figure 1.5: The CTIS raw data set. The central, white image is the zero order image that determines the spatial resolution of the instrument. The higher order projections are replications of the zero order image with dispersion.

The color in the data set shows explicitly the overlap of the spatial and spectral information as a spectral smear of the central (zero order) image. The physical layout and mathematical model of the CTIS and its reconstruction algorithms will be discussed in the next chapter.

Tunability and dissertation goals

One of the key elements of the CTIS is its grating disperser. It is the component that controls the number of projections (diffraction orders), the dispersion, and the signal to noise ratio by its aperture and spectral efficiencies. It is also the component that, if it

could be changed, would make the CTIS 'tunable' in the sense that it could be used to change the instrument's configuration parameters. Specifically, tunability allows the grating period (affecting the dispersion and spectral resolution), the number of orders (which is the number of projections from which to reconstruct), and to some degree the spectral efficiencies of the orders (which allows for possible *optical* data cube filtering) to be changed. The *spatial* resolution of the instrument *cannot* be changed with a tunable disperser. The spatial resolution is set by the field stop, FPA pixel sizes, and the CTIS lenses. Using zoom lenses in the CTIS allows both the spatial and spectral resolution of the CTIS to be altered, but without a tunable disperser the number of diffraction orders and their spectral efficiencies cannot be changed. The combination of zoom lenses and a tunable disperser allows for the full tunability of all CTIS configuration parameters. The instrument built for this project has this combination; however, to limit the scope of the project only the tunability of the disperser is discussed.

Current CTIS dispersers are phase Computer Generated Holographic (CGH) diffraction gratings used in transmission. After the phase values of these CGHs are designed, they are fabricated by etching a surface relief pattern into some material. In the visible wavelength region, this material is polymethyl methacrylate (PMMA) plastic, and in the MWIR either silicon or gallium arsenide (GaAs) are used. Tunability requires that the CGH phases can be changed with some reasonable speed, and liquid crystal spatial light modulators (LC SLM) allow for fairly rapid, programmable tunability of the phase values. Raytheon Network Centric Systems (RNCS) in Lexington, Massachusetts, has developed a liquid crystal spatial light modulator that is a good candidate for use as a

tunable CTIS disperser. RNCS refers to their device as an Optical Phased Array[7] (OPA) in direct analogy with microwave phased arrays used for beam steering. Raytheon has been kind enough to lend their demonstration OPA for this project.

This dissertation is an evolutionary work building on the groundwork laid by others, most notably M. Descour[8] and C. Volin[9]. The theoretical basis of CTIS operation, calibration, reconstruction, and disperser design are found in their dissertations and are summarized in Chapters 2 and 3 for completeness. There are two goals for this work: (1) to integrate a liquid crystal spatial light modulator into the CTIS and characterize its performance as a tunable CTIS disperser, and (2) to research the effects of tunability by implementing multiple CTIS configurations using the LC SLM. The details of integrating the Raytheon OPA into the CTIS require a modification of current CGH design algorithms. A summary of the current CGH design algorithm is found in Chapter 3, and the application of the CGH design algorithm with the OPA, as well as the OPA's physical layout and grating performance, are contained in Chapter 4.

Chapter 5 covers the implementation of tunability. Specifically, the effects of three configurations on the reconstruction speed, signal to noise ratio of the diffracted orders, and especially the spectral resolution limit of the instrument are explored.

CHAPTER 2 – CTIS INSTRUMENT

CTIS optical layout

The CTIS consists of three lens elements, a field stop, a dispersing element, and a large format focal plane array (FPA). A schematic of the CTIS is shown in Figure 2.1(a) and a picture of the instrument used for this project is shown in Figure 2.1(b).

(a)   (b)
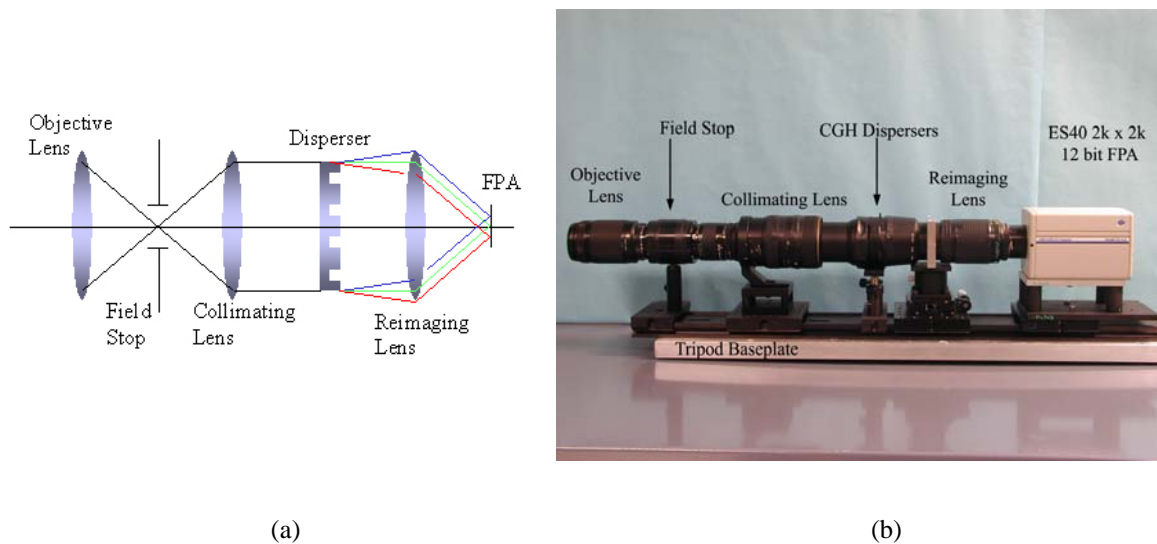
Figure 2.1: (a) Schematic drawing of the CTIS and (b) a photo of the visible CTIS instrument.

The optical train in Figure 2.1 proceeds from left to right. The objective lens forms an image of a scene in the plane of the field stop, which determines the object space field of view of the instrument. The field stop and its contents are reimaged to the FPA by the collimating/reimaging telescope with some minification. This image is termed the zero-order image. The higher diffraction orders of the disperser serve to replicate this zero-order image in a rectangular array. Each replication has some amount of dispersion that is determined by the disperser's grating period and the reimaging lens focal length. This

forms the raw data set that the CTIS collects, as shown in Figure 1.5 in the previous

chapter. This data set contains the 2D *projections* of the 3D object cube from which the

3D data cube (which is the best estimate of the continuous, real object cube) is

reconstructed using techniques similar to limited angle tomography.

CTIS mathematical model

As Volin[9] states in his dissertation "… it is not my intention to re-derive the

system model …." Descour[8] carries through the derivation of the system model, and

Volin summarizes the results clearly and concisely while describing the current state-of-

the-art reconstruction methods. These dissertations together provide a more than

adequate mathematical basis for CTIS operation, and the highlights from these

dissertations are included here for completeness. The structure of the description

presented here follows Descour, starting with the continuous-to-continuous model

describing the frequency space representation of the object cube and the missing cones,

then describes the more accurate continuous-to-discrete model taking into account

quantization and pixellation in the measurements, and finally ends with the discrete-to-

discrete model that must be used to reconstruct the data cube using a discrete set of basis

functions (voxels).

Continuous-to-continuous model

The object cube, f($\mathbf{r}$,λ), is a continuous function of its arguments (where

$\mathbf{r} = x\hat{\mathbf{x}} + y\hat{\mathbf{y}}$). The continuous model of the CTIS imaging system is given by

equation (2.14) in Descour, and is rewritten by Volin as:

$$g(\mathbf{r}') = \sum_{j=1}^{\infty} \iiint f(\mathbf{r}, \lambda) \, t_j(\lambda) \, \delta(\mathbf{r} - \mu\mathbf{r}' + \mu m_j \lambda \hat{\mathbf{r}}_j') \, d^2\mathbf{r} \, d\lambda + n(\mathbf{r}') \qquad (2.1)$$

The integral term represents each diffraction order and the sum index, j, is over the

multiple diffraction orders. The vector $\mathbf{r}'$ denotes the spatial coordinate vector of the

image, whereas the vector $\mathbf{r}$ is the spatial coordinate vector of the object in the plane of

the CTIS *field stop*. The function $t_j(\lambda)$ denotes the spectral transmittance of the CTIS for

the jth order, which is dominated by the spectral efficiency of the disperser and the

camera's spectral response (assuming that the optics are well corrected for aberrations).

The scalar μ is the magnification between the image and the object in the field stop, and

m_j is the dispersion factor for the jth order *in the direction of* $\hat{\mathbf{r}}_j'$. The function n($\mathbf{r}'$)

denotes additive noise. Equation (2.1) was formed by integration along idealized Dirac-δ

lines as part of the continuous, tomographic model. For simplicity in the frequency

domain representation, it is assumed that the spectral transmittance of the instrument is

unity and that there is zero noise. Each particular diffraction order is represented in the

frequency domain by taking the Fourier transform of the integral terms within the

summation of equation (2.1):

$$G_j(\rho') = F\left(\xi/\mu, \eta/\mu, m_j \rho' \cdot \hat{\mathbf{r}}_j'\right) \qquad (2.2)$$

Each diffraction order defines a plane in the 3D frequency space representation of the

object cube, and each plane must pass through the origin in frequency space in what is

known as the *central slice theorem*. The central slice theorem indicates that additional

data can be obtained only by changing the rotation angle ($\hat{r}_j{}'$) and inclination (m_j) of each

plane. Figure 2.2(a) shows a representation of $G(\rho')$ in frequency space coordinates.

(a)   (b)
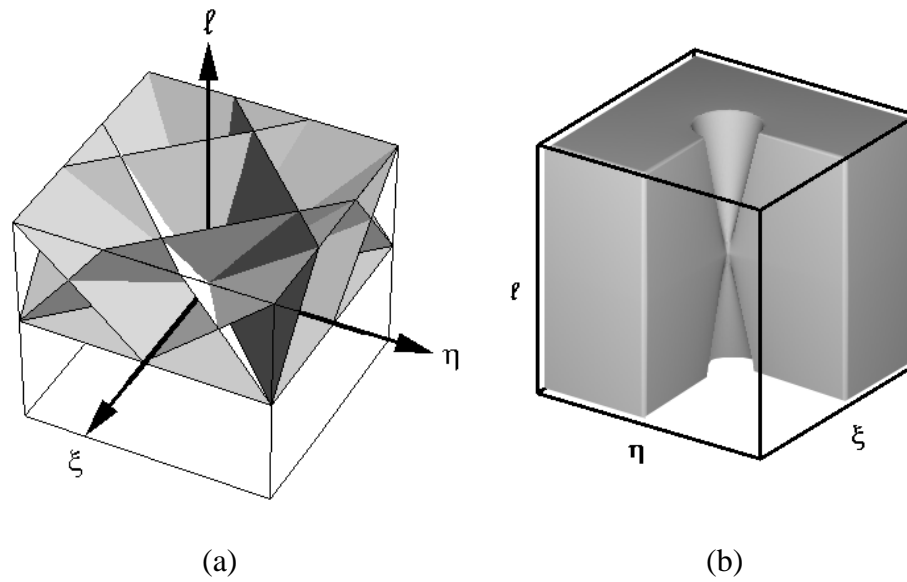
Figure 2.2: Depiction of the data cube with (a) a finite number of planes and (b) an infinite number of

planes that are possible with the CTIS illustrating the missing cones.[9]

The plane with zero inclination represents the zero order, as it has no dispersion. There is

an inherent limitation on the inclination angle of the data in frequency space that is

collected by the CTIS, namely, the spatial extent of the FPA. Taking data of an infinite

number of planes (CTIS diffraction orders) over all the possible angles (rotation of the

planes about the *l* axis) and inclinations sweeps out the range of viable data space shown

in Figure 2.2(b). The limitation on the inclination angle causes there to be a region where information at high spectral/low spatial frequencies is lost in 'cones' of empty space. These are the so-called missing cones, one above and below the ξ–η plane. Near infinite dispersion across the FPA would be needed to fill in data along the l axis.

Following Volin, a more accurate rendering of the situation is obtained by carrying through the transmittance term, $t_j(\lambda)$:

$$G_j(\rho') = F\left(\xi/\mu, \eta/\mu, m_j\boldsymbol{\rho}'\cdot\hat{\mathbf{r}}_j'\right) * * * \delta(\xi/\mu)\delta(\eta/\mu)T_j(\boldsymbol{\rho}'\cdot\hat{\mathbf{r}}_j') \qquad (2.3)$$

Typically the transmittance of the instrument is dominated by the spectral diffraction efficiency of the disperser, which is designed to be spectrally flat. This transforms to a delta function in wavelength frequency space and equation (2.3) reduces to the form of equation (2.2). What equation (2.3) indicates, however, is that altering the diffraction efficiencies of the disperser will change the performance of the CTIS *and* that the information may *possibly be spread into the missing cone* by designing the disperser with some spectral efficiency modulation. Volin has explored methods to fill in the missing cone using a priori estimates and SVD. Modulating the disperser efficiencies presents an additional, albeit difficult, possibility for extrapolating information into the missing cone, but is beyond the scope of this work.

Continuous-to-discrete model

The most accurate model of the CTIS is the continuous-to-discrete model where the object cube is represented as a continuous function of its arguments. However, the

data that the CTIS collects is discretized due to the spatial extent of the FPA pixels, A/D

conversion for storing the image digitally, and the use of a finite number of diffraction

orders. This is written as:

$$g_i = \sum_{j=1}^{N} \iiint f(\mathbf{r}, \lambda) \, t_j(\lambda) \, h_i(\mathbf{r} - \mu\mathbf{r}' + \mu m_j \lambda \hat{\mathbf{r}}_j') \, d^2\mathbf{r} \, d\lambda + n_i \qquad (2.4)$$

The image is no longer a function of a continuous 2D spatial vector but rather is a

collection of a finite number of locations (pixels) shown by the subscript i. The

summation is over a finite number of diffraction orders, and the A/D conversion is

lumped into the noise term as quantization noise. The integration manifolds are no

longer the idealized line Dirac-δ functions of the continuous model and are now

represented by the linear operator h, denoting either the PSF of the instrument if the CTIS

is viewed as a linear imaging system, or the probability of a photon arising from $(\mathbf{r}, \lambda)$

being recorded at pixel i, if statistical optics is preferred. The discrete nature of the

collected data prevents the data from fully sampling the volume of frequency space

shown in Figure 2.2(b), and in actuality the situation is better represented by Figure

2.2(a) where there are areas of the object cube that are unsampled in frequency space.

Thus a unique reconstruction of an object cube from the CTIS projections cannot be

formed. However, the use of non-linear reconstructions techniques utilizing a priori

information about the object cube results in a good approximation of the object cube.

<u>Discrete-to-discrete model</u>

Although the continuous-to-discrete mathematical model is the most accurate, a continuous object cube cannot be reconstructed. Thus the term *data cube* is used to denote the reconstructions. The data cube is an approximation of the object cube using a discrete set of basis functions – voxels. The imaging equation now becomes:

$$\mathbf{g} = \mathbf{H}\hat{\mathbf{f}} + \mathbf{n} \qquad\qquad (2.5)$$

where $\mathbf{g}$ is an M x 1 image vector, $\mathbf{H}$ is the M x N system matrix, $\hat{\mathbf{f}}$ is an N x 1 object vector, and $\mathbf{n}$ is an M x 1 noise vector. The 2D CTIS image, $\mathbf{g}$, and 3D data cube, $\hat{\mathbf{f}}$, are arbitrarily rasterized into vectors for matrix multiplication. Equation (2.5) is termed the *forward* imaging problem. The goal is to invert this equation to obtain the data cube $\hat{\mathbf{f}}$.

Reconstruction techniques

The direct inversion of equation (2.5) using a pseudoinverse is impractical due to the size of the system matrix. As an example, consider the 2048 x 2048 FPA used for this project with a zero order image size of 100 x 100 pixels and 30 reconstruction bands. A single voxel maps to 2048 x 2048 FPA pixels. There are 100*100*30 voxels, giving a system matrix size of roughly $1.2 *10^{12}$ elements. This is somewhat misleading, as for each voxel most of the 2048 x 2048 pixels do not contain information about the voxel spread function and can be stored sparsely. Even so, direct inversion is intractable and consequently the CTIS reconstruction techniques are all iterative using sparse matrix

storage and multiplication in the algorithms. Descour employed the additive techniques

of the Landweber algorithm and the Maximum Likelihood approach, which assume

signal *independent* and signal *dependent* noise, respectively, as well as the multiplicative

technique of Expectation Maximization (EM), which also assumes signal dependent

noise obeying Poisson statistics. Descour found the multiplicative technique to be the

most satisfactory as it has very rapid, non-linear convergence. One of the drawbacks of

EM is that it tends to introduce high-frequency artifacts in the reconstruction. Therefore,

Volin added a Multiplicative Algebraic Reconstruction Technique (MART) to the CTIS

reconstruction library, which assumes a Gaussian noise distribution and tends to reduce

the high-frequency spatial artifacts introduced by EM. These are the current workhorse

algorithms used in CTIS reconstructions. The iterative equation for EM is listed below in

equation (2.6) and the equation for MART is listed in equation (2.7), where k is the

iteration index and the starting point is typically a uniform initial estimate of the data

cube.

(EM)
$$\hat{f}_n^{k+1} = \frac{\hat{f}_n^{k}}{\sum\limits_{m'=1}^{M} H_{m'n}} \sum_{m=1}^{M} H_{mn} \frac{g_m}{\left(\mathbf{H}\hat{\mathbf{f}}^{k}\right)_m} \tag{2.6}$$

(MART)
$$\hat{f}_n^{k+1} = \hat{f}_n^{k} \frac{\left(\mathbf{H}^{T}\mathbf{g}\right)_n}{\left(\mathbf{H}^{T}\mathbf{H}\hat{\mathbf{f}}^{k}\right)_n} \tag{2.7}$$

Calibration

The CTIS reconstruction algorithms are dependent on the system matrix **H**. The system matrix can be determined computationally, experimentally, or by a hybrid of both methods. Computationally calculating the system matrix is faster than determining the matrix in the lab, however it depends on accurate knowledge of the optics and their coatings, the disperser along with its fabrication errors, the camera response and pixel fill factor, scattering, and numerous other factors of reality that complicate the computation. On the other hand, experimentally characterizing the system matrix is exhaustive. To date, a hybrid approach has been taken utilizing the shift invariance of the CTIS.

Figure 2.3 shows the hardware that is used for the calibration. A DC powered tungsten-halogen lamp is imaged to the entrance slit of a TRIAX 180 monochromator manufactured by Jobin-Yvon. The light is collimated and directed to one of three blazed reflection gratings mounted on a rotatable turret. The gratings are blazed at 1200 lines/mm, 600 lines/mm, and 300 lines/mm for use at visible, SWIR, and MWIR wavelengths, respectively. The collimated, dispersed light is then focused to the exit slit, which samples the dispersed spectra. The wavelength exiting the monochromator is selected by rotating the grating, and the entrance and exit slits can be opened and closed to control the bandwidth and irradiant intensity of the exiting light. The spectral content of the light exiting the monochromator is given by a convolution of the entrance and exit slits and should thus be triangular, which was confirmed upon observation with a reference spectrometer. The exiting light is then focused to the entrance face of a fiber by a microscope objective. The microscope objective is used because it has been found
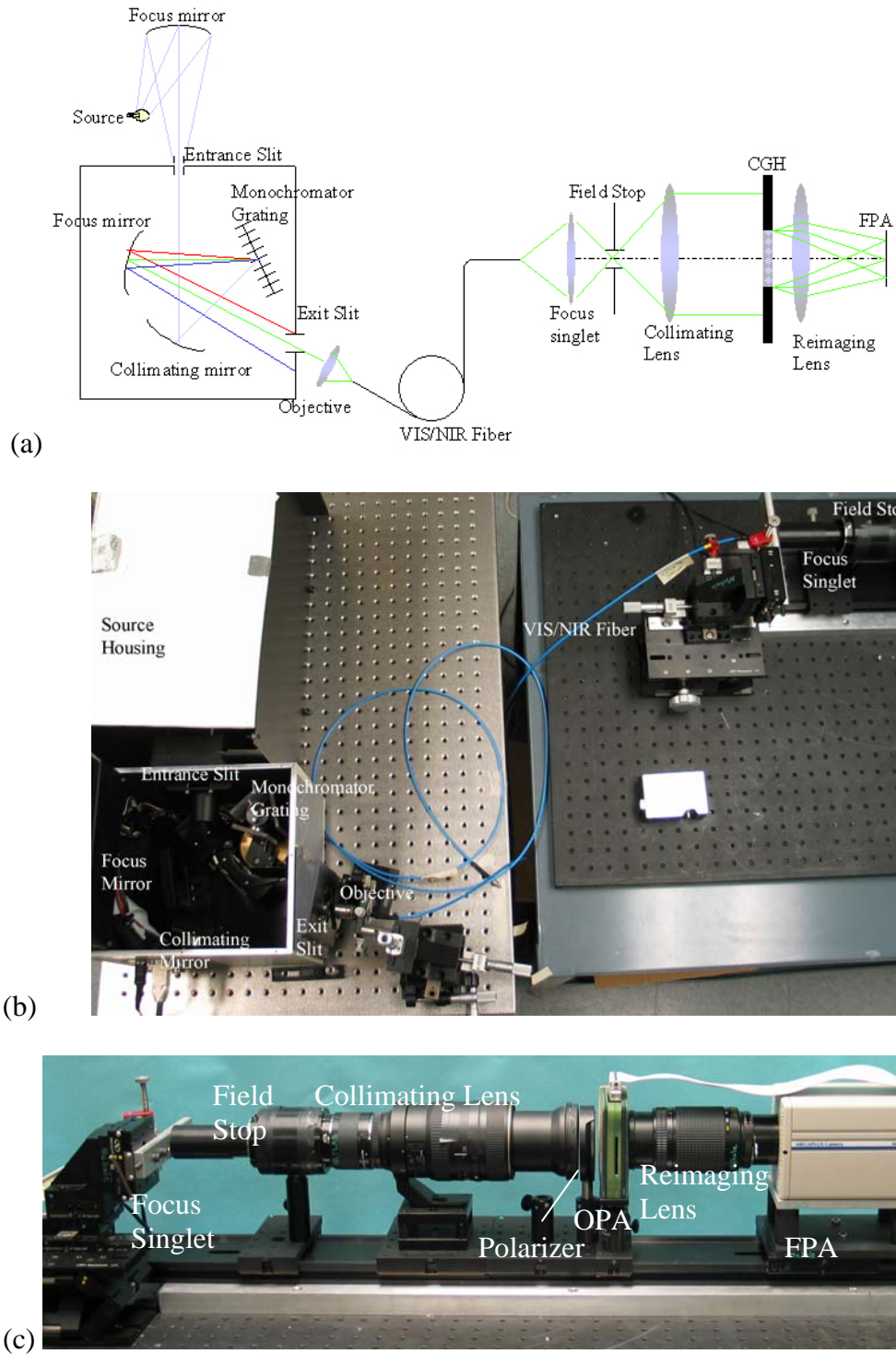
Figure 2.3: (a) A schematic of the experimental calibration, (b) a picture of the calibration apparatus (left half of schematic) and (c) the CTIS (right half of schematic).

to couple the most light to the fiber compared with spherical or off-axis parabolic mirrors, which would be necessary when working in the MWIR. The exit face of the wave guide is mounted using xyz translation stages and is imaged to the CTIS field stop. If the image of the fiber's exit face is small enough, a single pixel on the CTIS FPA will be illuminated in the zero order image and a spectral/spatial point source has effectively been placed in the plane of the field stop. The point source can be translated to different locations in the field stop using the stages, and a quasi-monochromatic wavelength can be selected using the monochromator. This point source serves to illuminate a single *voxel* of the data cube, and the CTIS calibration records the response of the instrument to individual voxels. An image acquired with this set up is therefore referred to as a *voxel spread function* (VSF), which is the 3D counterpart to a *point spread function* (PSF). A full experimental calibration requires acquiring a VSF for every voxel in the data cube.

Figure 2.4 schematically shows the relationship between the system matrix and individual voxels. In Figure 2.4(a), the center voxel is imaged through the system by imaging the exit face of the wave guide on-axis to the field stop with the monochromator selecting the wavelength that corresponds to the center band. Since only one voxel contains light, the vector **f** will have only one non-zero value, shown at location n'. The resulting image replicates the n' column of the system matrix multiplied by the constant $f_{n'}$ (the value of the n' element of **f**). This calibration image constitutes the VSF of the n' voxel. A reference photometer, manufactured by United Detector Technologies (UDT), is used to normalize the values to the appropriate units. The UDT photometer has a responsivity in units of Amps/Watt for the optical power incident on its detector, which is

the optical power in Watts of the voxel. The image data is quantized as the unitless

numbers of analog-to-digital units (ADUs), and the integration time of the camera is used

along with the reference photometer readings to normalize the units of the system matrix

H to ADU/Joule.

$$\begin{pmatrix} g_1 \\ g_2 \\ \cdots \\ \cdots \\ g_{M-1} \\ g_M \end{pmatrix}_{x'} = \begin{pmatrix} h_{1,1} & h_{1,2} & \cdots & \begin{pmatrix} h_{1,x'} \\ h_{2,x'} \\ \cdots \\ \cdots \\ h_{M-1,x'} \\ h_{M,x'} \end{pmatrix} & \cdots & h_{1,N} \\ h_{2,1} & \cdots & \cdots & & \cdots & \cdots \\ \cdots & \cdots & \cdots & & \cdots & \cdots \\ \cdots & \cdots & \cdots & & \cdots & \cdots \\ \cdots & \cdots & \cdots & & \cdots & \cdots \\ h_{M,1} & \cdots & \cdots & & \cdots & h_{M,N} \end{pmatrix} \begin{pmatrix} 0 \\ 0 \\ \cdots \\ f_{x'} \\ \cdots \\ 0 \\ 0 \end{pmatrix}$$

If $f_{x'} = 1$, then $\mathbf{g}_{x'} = \mathbf{h}_{x'}$.

$$\begin{pmatrix} \cdots & \cdots & \cdots & \begin{pmatrix} h_{1,x'} \\ h_{2,x'} \\ \cdots \\ \cdots \\ h_{M-1,x'} \\ h_{M,x'} \end{pmatrix} \begin{pmatrix} h_{1,x'+1} \\ h_{2,x'+1} \\ \cdots \\ \cdots \\ h_{M-1,x'+1} \\ h_{M,x'+1} \end{pmatrix} \cdots \begin{pmatrix} h_{1,x'+l-1} \\ h_{2,x'+l-1} \\ \cdots \\ \cdots \\ h_{M-1,x'+l-1} \\ h_{M,x'+l-1} \end{pmatrix} & \cdots & \cdots \\ \end{pmatrix}$$
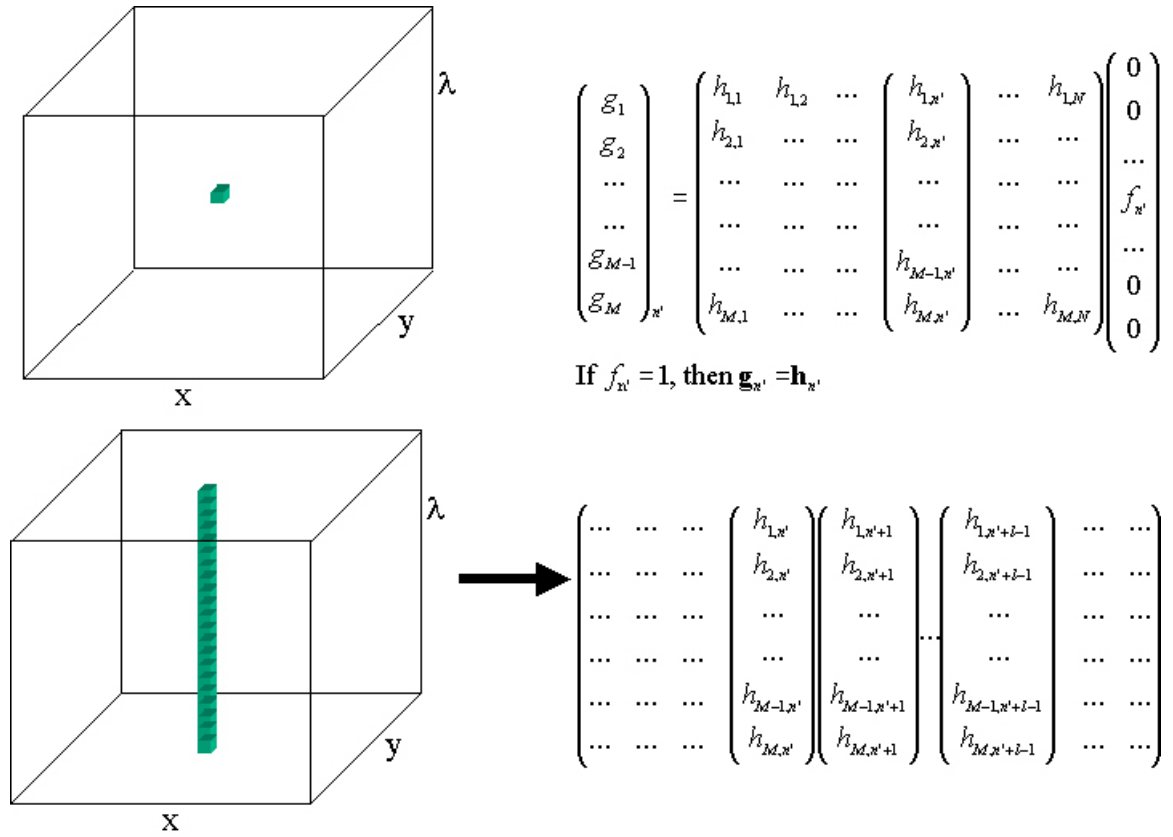
Figure 2.4: (a) Illustration of the formulation of a column of the system matrix corresponding to a single VSF. (b) The set of data taken for the hybrid experimental/computational calibration that samples a column of the data cube voxels and forms a set of columns in the system matrix.

The hybrid calibration procedure only requires experimental determination of the

VSFs from a subset of the voxels of the data cube. The CTIS is assumed to be shift

invariant for each wavelength, therefore the VSFs for each CTIS band need only be experimentally acquired for a single spatial position, typically on axis, as shown in Figure 2.4(b). The remaining columns of the system matrix are filled in by appropriately shifting the VSFs of the experimental data appropriately.

Shift invariance is generally true of the CTIS instrument, however, each individual instrument should be tested for shift invariance due to variations in the optics and alignment. The calibration images for the CTIS used in this project have a correlation of just over 98% from the center to the corner of the field stop. Figure 2.5 shows two images superimposed, one with the point source located on axis and the other with the point source located near the upper right corner of the field stop (the box at the center denotes the spatial extent of the field stop). The images have been scaled to show the order locations.
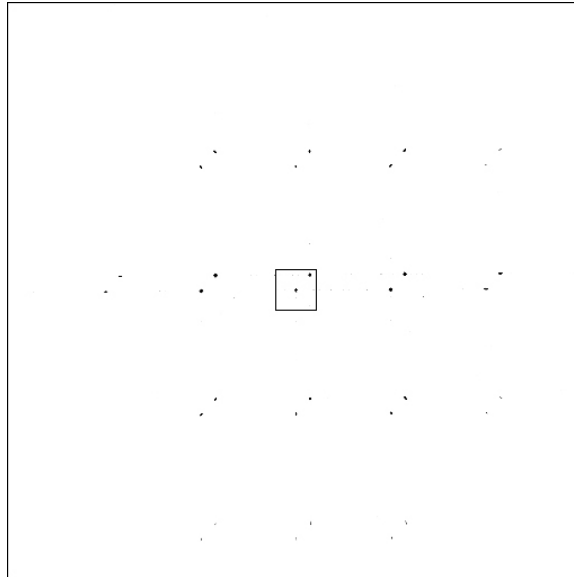
Figure 2.5: Comparison of VSFs for locations at the center and corner of the field stop.

In reality, it is difficult to image a point source in the plane of the field stop to the 7.4 um wide pixels of the FPA used for this project, and the image has some spread.
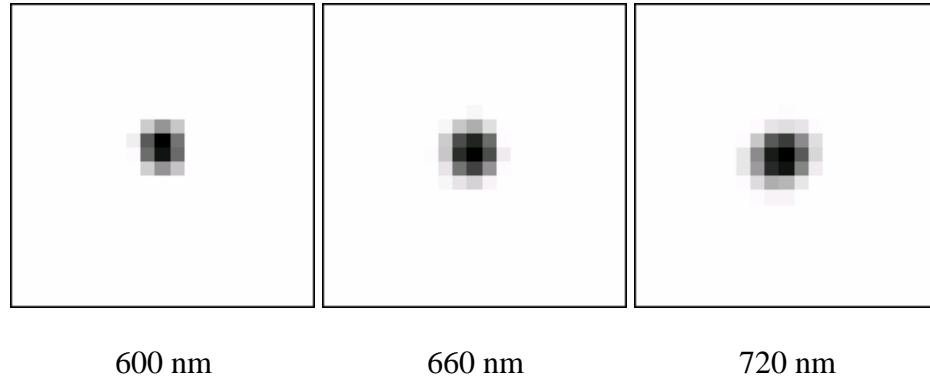
600 nm   660 nm   720 nm

Figure 2.6: Zero order images at 600 nm, 660 nm, and 720 nm.

Figure 2.6 shows magnified versions of the zero order for calibration images taken at 600 nm, 660 nm, and 720 nm. The fiber face cannot image to a single pixel, partly because the fiber needs to be large to get sufficient signal, and partly because the CTIS itself consists of off-the-shelf optics and a borrowed OPA that have not been designed to work together for optimal imaging. The blur increases with wavelength because of chromatic aberration in the singlet that is used to image the fiber face to the plane of the field stop. The optimal focus of the singlet is set at 600 nm, the lowest wavelength, to increase the signal as much as possible at this end of the wavelength band. The effect of using 'blurred' point sources to determine the system matrix is to blur the reconstructions and to degrade the spectral resolution of the instrument, which will be explored in more depth in Chapter 5.

CHAPTER 3 – CTIS DISPERSER DESIGN

The first dispersive elements used in the CTIS by Descour[8] were a trio of

sinusoidal phase gratings crossed at 60^0 to effect a hexagonal pattern of projections on the

FPA. The far-field efficiencies of the diffracted orders of a sinusoidal phase grating are

calculated by the Fourier transform of the grating's transmittance[10] and there is little

freedom to tailor these efficiencies to a desired pattern. The gratings were high

efficiency, but most of that efficiency was not fanned out evenly into the desired

diffraction orders. Consequently, it was difficult to fully utilize the dynamic range of the

camera and the distribution of light caused some of the desired orders to have low SNR.

Phase computer generated holograms (CGHs) allow for greater control over the

efficiencies of the orders by providing the degrees of freedom needed in the CGH plane

to optimize the coupling of the light into the desired diffraction orders.

There are many design algorithms which calculate the phase profile of a CGH

needed for a particular far-field irradiance pattern including simulated annealing,[11]

conjugate gradient,[12] downhill simplex,[13] and Gerchberg-Saxton.[14] However, these are

typically implemented as monochromatic design algorithms. While it is possible to use a

polychromatic merit function to optimize the design for a wavelength band of interest

using these algorithms, the convergence to an optimum solution requires a large set of

individual designs.[15] Volin has developed a linear inversion technique[16,9] for the design

of polychromatic CGHs, and this is the state of the art design algorithm used for CTIS

dispersers. For completeness, the development of the technique is repeated here roughly

following the pattern that can be found in Volin, but with the design results presented for

the visible wavelength region rather than the MWIR. Modifications to the technique are

required for application with the Raytheon optical phased array (OPA) and are described

in the next chapter. The results of a few example CGH designs are presented at the end

of this chapter.

Polychromatic SVD CGH design technique

The design goal is to determine the correct CGH phases for a given desired

amplitude pattern in the far-field. It is typically assumed that plane waves are incident on

the CGH, and this is the case with the CTIS. The following description comes in four

parts: (1) the mathematical model of the CGH, (2) linearity, (3) singular value

decomposition, and (4) linear inversion.

<u>Mathematical model of the CGH</u>

The CTIS CGH consists of a single unit cell of individual phase elements referred

to as either 'phasels' or CGH pixels. The unit cell typically contains 16 x 16, 10 x 10, or

8 x 8 phasels, and this cell is identically replicated across the entire clear aperture of the

CGH.[17] The phase delay of a particular phasel is given by:

$$\Phi_\lambda = \frac{2\pi(n_\lambda - 1)d}{\lambda} \tag{3.1}$$

where n_λ is the material index at wavelength λ, and d is the physical etch depth in the material. The transmission of the CGH is given by:

$$t(\xi,\eta) \propto \sum_{m,n=0}^{q-1} e^{j\Phi_{mn}} rect\left(\frac{\xi - m\xi_c}{\xi_c}\right) rect\left(\frac{\eta - n\eta_c}{\eta_c}\right) ** \sum_{k,l=-\infty}^{\infty} e^{j\Phi_{mn}} \delta\left(\frac{\xi - kq\xi_c}{q\xi_c}\right) \delta\left(\frac{\eta - lq\eta_c}{q\eta_c}\right) \tag{3.2}$$

where q is the number of phasels in each direction, ξ_c and η_c are the physical width and height of each phasel, and ** denotes a two-dimensional convolution.

(a)   (b)
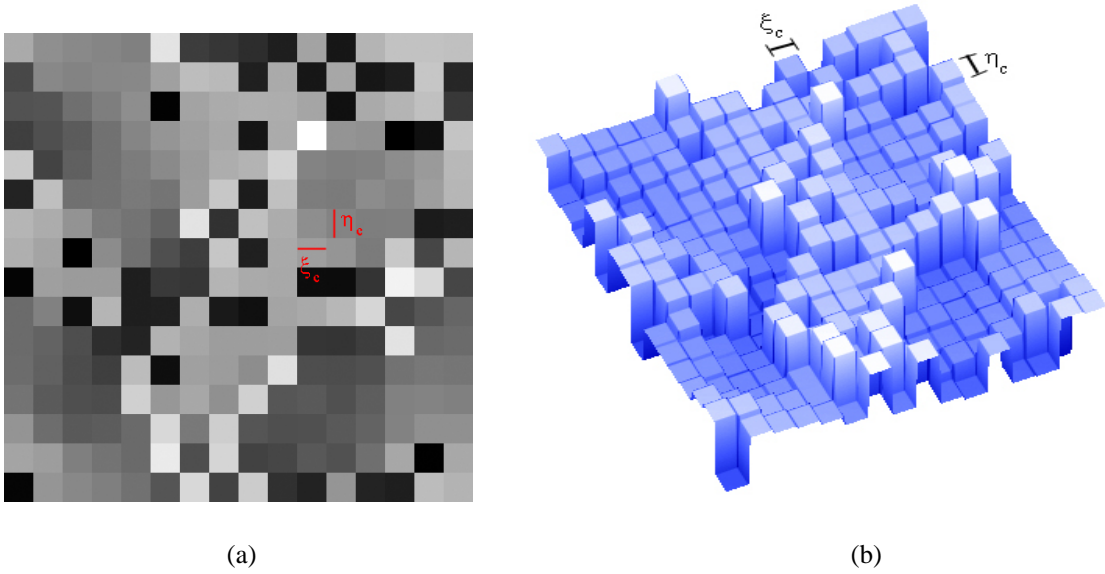
Figure 3.1: (a) Grayscale representation of the CGH unit cell as well as (b) a surface plot representation of a q = 16 etched CGH unit cell.

The first term of equation (3.2) denotes the unit cell of CGH phasels, and the second is the replication term. Figure 3.1(a) shows a grayscale representation of the unit cell of

equation (3.2) with q = 16. Figure 3.1(b) shows a surface plot of what the unit cell of the resulting etched CGH would look like.

The far-field complex amplitude of plane waves encountering this CGH followed by a lens of focal length f is given by the Fourier transform of equation (3.2) according to Fraunhofer diffraction. The transform is evaluated at $\xi = x/\lambda f$ and $\eta = y/\lambda f$ to give:

$$\mathrm{E}(x, y) \propto \mathfrak{I}_2\left\{t(\xi, \eta)\right\}\big|_{\xi=\frac{x}{\lambda f}, \xi=\frac{x}{\lambda f}}$$

$$\mathrm{E}(x, y) \propto \sum_{m,n=0}^{q-1} e^{j\Phi_{mn}} e^{\frac{-j2\pi m \xi_c x}{\lambda f}} e^{\frac{-j2\pi n \eta_c y}{\lambda f}} \times \sum_{k,l=-\infty}^{\infty} \delta\left(x - \frac{k\lambda f}{q\xi_c}\right)\delta\left(y - \frac{l\lambda f}{q\eta_c}\right) \times \mathrm{sinc}\left(\frac{\xi_c x}{\lambda f}\right)\mathrm{sinc}\left(\frac{\eta_c y}{\lambda f}\right)$$

(3.3)

where $\mathrm{sinc}(x) = \sin(\pi x)/\pi x$. Equation (3.3) is nonzero only at the delta function locations, which are given by $x_k = k\lambda f / q\xi_c$ and $y_l = l\lambda f / q\eta_c$, so a simplified version of equation (3.3) can be rewritten as:

$$\mathrm{E}(x_k, y_l) \propto \sum_{m,n=0}^{q-1} e^{j\Phi_{mn}} e^{\frac{-j2\pi mk}{q}} e^{\frac{-j2\pi nl}{q}} \times \mathrm{sinc}\left(\frac{k}{q}\right)\mathrm{sinc}\left(\frac{l}{q}\right)$$

$$k = -\infty, +\infty$$

$$l = -\infty, +\infty$$

(3.4)

Notice that after scaling by the incident wave amplitude (or assuming the incident light is unit amplitude), this is a modified form of the Discrete Fourier Transform (DFT):

$$U_{kl} = \frac{1}{\sqrt{ML}} \sum_{m=-M/2}^{M/2-1} \sum_{n=-N/2}^{N/2-1} u_{mn} e^{-j2\pi\left(\frac{km}{M} + \frac{ln}{N}\right)}$$

(3.5)

and is even more closely related to the form of the DFT found in computing software such as IDL:

$$U_{kl} = \frac{1}{MN} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} u_{mn} e^{-j2\pi\left(\frac{km}{M} + \frac{ln}{N}\right)}$$

(3.6)

The DFT then gives the *exact* Fourier transform of the phase pattern when $u_{mn} = e^{-j\Phi_{mn}}$ and

after multiplication by the correct sinc function. Note that the first lobe of the sinc

function contains 2q orders and its zeros occur every q orders; i.e. the zeros are at n*q

where n is an integer ranging from negative to positive infinity. Also note that from

equation (3.4) the far-field pattern repeats exactly (except for the sinc function) every q

orders. The amplitudes of the first 2q orders resulting from the phase profile of Figure

3.1 are shown below in Figure 3.2.

(a)   (b)
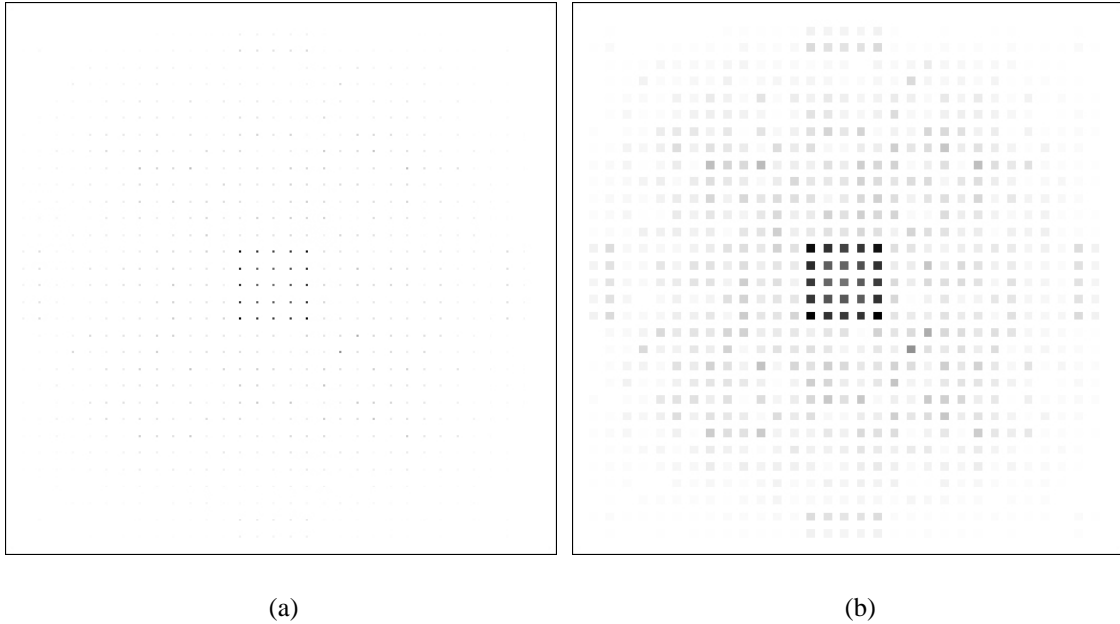
Figure 3.2: (a) The diffraction order amplitudes (in reverse contrast) of the phase profile in Figure 3.1.

(b) The very same image with the orders depicted as pixels for better visibility. The first 2q orders are

shown, with the edges corresponding to the first zero of the sinc function.

<u>Linearity</u>

The mapping from the CGH plane to the far-field plane is a finite, discrete-to-discrete mapping. The Fourier transform is a linear operation between the CGH *transmission* and the *electric fields* in the Fraunhofer region. However, the CGH *phases* need to be related to the electric field *amplitudes*, and there are two non-linear operations conducted along the mathematical train to relate those quantities: (1) the transmission of the CGH is the exponential of the phase delay of the material times its amplitude transmission (which is 1 for the ideal phase CGH) - $t(x_m, y_n) = T_{mn} e^{-j\Phi_{mn}}$, and (2) the amplitudes are the absolute value of the complex electric fields $A_{kl} = |E_{kl}|$. For the model it is reasonable to assume that the CGH is completely transmitting, $T = 1$. The non-linear steps can be illustrated as:

$$\Phi_{mn} \underset{non-linear}{\to} e^{i\Phi_{mn}} \underset{linear}{\overset{\mathcal{F}\{\}}{\Longleftrightarrow}} E_{kl} \underset{non-linear}{\to} \left|E_{kl}\right| = A_{kl} \tag{3.7}$$

The far-field amplitudes are certainly not linearly related to the CGH phases, but it can be shown that small *changes* in CGH phase *are* linearly related to changes in the far-field amplitudes, i.e.:

$$\Delta A_{kl} = \sum_{n=0}^{q-1} \sum_{m=0}^{q-1} \mathcal{H}_{\Phi,klmn} \Delta \Phi_{mn} \tag{3.8}$$

where $\mathcal{H}_{\Phi,klmn}$ is the linear operator, called the sensitivity matrix, and the first subscript denotes that the values depend on the current CGH phase values. Linearity is shown by a perturbing a single CGH phasel by a small phase change, $\Phi'_{mn} = \Phi_{mn} + \Delta\Phi_{mn}$. The approximate transmission of the CGH at that phasel is:

$$e^{j\Phi'_{mn}} \approx e^{j\Phi_{mn}} + j\Delta\Phi_{mn}e^{j\Phi_{mn}} \tag{3.9}$$

Notice that the transmission of the perturbed phasel is approximated by the original

transmission plus a perturbation term that is linear with $\Delta\Phi_{mn}$. Also, a perturbation at a

different phasel will be separately additive. For a given plane wave impinging on the

CGH, $\mathbf{E} = |E|e^{j\Psi}$, the resulting complex electric field in the Fraunhofer region is:

$$\mathrm{E}'(x_k, y_l) = \mathrm{E}(x_k, y_l) + j\Delta\Phi_{mn}e^{j\Phi_{mn}-\Psi}e^{\frac{-j2\pi n\xi_c x_k}{\lambda f}}e^{\frac{-j2\pi n\eta_c y_l}{\lambda f}} \tag{3.10}$$

Using $x_k = k\lambda f / q\xi_c$ and $y_l = l\lambda f / q\eta_c$ from equation (3.3), the discrete diffraction

orders are written as:

$$\mathrm{E}'(x_k, y_l) = \mathrm{E}(x_k, y_l) + j\Delta\Phi_{mn}e^{j\Phi_{mn}-\Psi}e^{\frac{-j2\pi nk}{q}}e^{\frac{-j2\pi ml}{q}} \tag{3.11}$$

$$\mathrm{E}'_{kl} = \mathrm{E}_{kl} + \Delta\Phi_{mn}e^{j\Theta_{klmn}-\Psi}$$

The change in complex field is then the product between the change in phase and three

complex exponential terms that have been lumped into Θ_{klmn}. The incident complex field

can be ordered into a vector and separated into real and imaginary parts as $\mathbf{E} = \mathbf{E_r} + j\mathbf{E_i}$

so that the complex electric field in the far-field can be written as:

$$\mathbf{E}' = \mathbf{E_r} + j\mathbf{E_i} + \Delta\Phi_{mn}e^{j\Theta_{klmn}} = \mathbf{E_r} + j\mathbf{E_i} + \mathbf{C} + j\mathbf{D} \tag{3.12}$$

The absolute value of the complex field gives the amplitudes of the orders:

$$\mathbf{A}' = |\mathbf{E}| = \left[(\mathbf{E_r} + \mathbf{C})^2 + (\mathbf{E_i} + \mathbf{D})^2\right]^{1/2} = \left[(\mathbf{E_r} + \mathbf{E_i}) + (\mathbf{C^2} + \mathbf{D^2}) + 2(\mathbf{E_r}\mathbf{C} + \mathbf{E_i}\mathbf{D})\right]^{1/2} \tag{3.13}$$

Equation (3.13) is expanded in a Taylor series about $\mathbf{C}$ and the change in the far-field

amplitudes to a first order approximation is:

$$\mathbf{A'} = \mathbf{A} + \left[\frac{\mathbf{E_r C}}{\mathbf{A}} + \frac{\mathbf{E_i D}}{\mathbf{A}} \right] + 2 \left[\frac{\mathbf{C^2 + D^2}}{\mathbf{A}} - \frac{(\mathbf{E_r C + E_i D})^2}{\mathbf{A^3}} \right] + \ldots \qquad (3.14)$$

$$\Delta\mathbf{A} = \mathbf{A'} \text{-} \mathbf{A} \cong \left[\frac{\mathbf{E_r}}{(\mathbf{E_r^2 + E_i^2})^{\frac{1}{2}}} \right] \mathbf{C} + \left[\frac{\mathbf{E_i}}{(\mathbf{E_r^2 + E_i^2})^{\frac{1}{2}}} \right] \mathbf{D} \qquad (3.15)$$

Linearity is violated by diffraction orders with a very small initial irradiance, which can be seen in the second order terms of equation (3.14) which have powers of the amplitude in the denominator. The linearity is extended in the algorithm by separately constraining the real and imaginary components of dim orders, effectively relating the phase to the complex electric field itself and simply not performing the second non-linear operation of equation (3.7). This bypasses the approximation due to taking only the first order term of the Taylor series. Note that this is possible only in the computer, and when the algorithm is used in feedback between the camera and the optical phased array (OPA) access is granted to only to the far-field amplitudes (as square roots of detected irradiances by the camera) and not the complex electric fields. Volin states that equation (3.15) is linear with an accuracy of 90% for phase perturbations of $\pi/7$ radians. Also, the change in amplitude for a particular order must be less than its current amplitude. Constraining the real and imaginary parts of low irradiance orders eliminates this restriction, and for the other orders the starting point must have some efficiency already coupled into orders that are desired to have non-zero efficiency. This will almost always be true when the starting point is a random phase profile.

Singular Value Decomposition

To this point, the mathematical description of the CGH and a linear relationship

between small phase changes and far-field amplitude changes has been established. The

inversion problem, however, is ill-posed because there are an unequal number of

measurements (diffraction amplitudes), M, and unknowns (CGH phasels), N. In this case

there are more measurements than unknowns, $M > N$, and the problem is said to be *over*

determined. A powerful technique for inversion of ill-posed, over determined linear

problems is singular value decomposition[18] (SVD). SVD generates a basis set for object

space (the CGH plane) and a basis set for image space (the far-field plane). For any

linear system operator $\mathcal{H}$ such that $\mathbf{g} = \mathcal{H}\mathbf{f}$, the operators $\mathcal{H}^\dagger\mathcal{H}$ and $\mathcal{H}\mathcal{H}^\dagger$ are Hermitian

and their eigenvalues $\{\mu_n\}$ are therefore real and positive and the eigenvectors ($\{\mathbf{v}_n\}$ and

$\{\mathbf{u}_n\}$, respectively) are real. The dagger notation indicates the adjoint of the matrix. The

eigenvectors from each set are orthonormal and related by:

$$\mathbf{u}_n = \frac{1}{\sqrt{\mu_n}}\mathcal{H}\mathbf{v}_n \tag{3.16}$$

The system matrix can be decomposed in terms of the eigenvectors and eigenvalues of

$\mathcal{H}^\dagger\mathcal{H}$ and $\mathcal{H}\mathcal{H}^\dagger$ as:

$$\mathcal{H} = UWV^t \tag{3.17}$$

The eigenvectors have been organized into matrices of column-oriented vectors $\mathbf{U}$ and $\mathbf{V}$,

and the square roots of the eigenvalues are placed along the diagonal of $\mathbf{W}$. The

eigenvectors are referred to as the *singular vectors* of $\mathcal{H}$ and the square roots of the

eigenvalues, $w_n = \sqrt{\mu_n}$, are referred to as the *singular values*. The matrix **V** contains

the object space singular vectors, also referred to as *phase perturbation modes*, and the

matrix **U** contains the image space singular vectors, which are also referred to as

diffraction order amplitude variation modes. The singular values rate the influence of a

particular object space singular vector on its corresponding image space singular vector

and are typically ordered from largest to smallest. The *rank* of the matrix, R, is equal to

the number of non-zero singular values. For an over determined equation, the rank is less

than or equal to the number of unknowns, $R \leq N$.

The exact inverse of the matrix exists only if the matrix is square, M = N, and

non-singular, R = N (i.e. there are no zero singular values).

$$\mathcal{H}^{-1} = \mathbf{V}\mathbf{W}^{-1}\mathbf{U}^t = \sum_{n=1}^{N} \frac{1}{w_n} \mathbf{v}_n \mathbf{u}_n^t \tag{3.18}$$

The exact inverse does not exist for singular, non-square matrices, but the Moore-Penrose

pseudoinverse can still be constructed:

$$\mathcal{H}^{+} = \mathbf{V}\tilde{\mathbf{W}}^{-1}\mathbf{U}^t = \sum_{n=1}^{R} \frac{1}{w_n} \mathbf{v}_n \mathbf{u}_n^t \tag{3.19}$$

where the summation is only over the rank. The tilde notation in equation (3.19) denotes

that the diagonal elements of the **W** matrix beyond the rank R (and up to N) are filled

with zeros rather than 1/zeros, as the equation would imply. The Moore-Penrose

pseudoinverse gives the least squares solution to the inversion problem.

Figure 3.3 shows an example decomposition of the CGH of Figure 3.1.

(a)

(b)

(c)

(d)
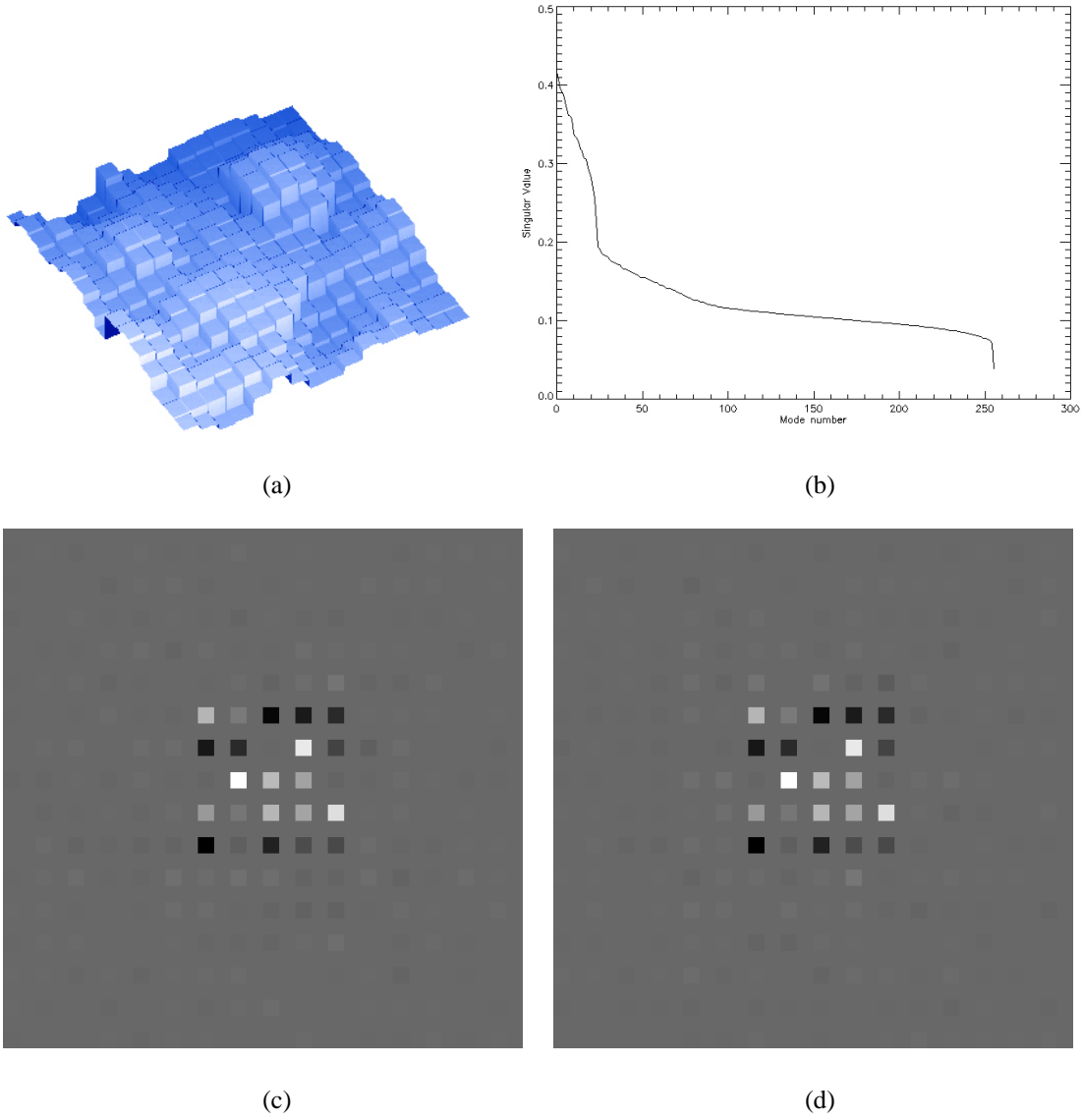
Figure 3.3: (a) The first phase perturbation mode (v_1), (b) the singular values (w), (c) the first diffraction order amplitude variation mode (u_1) with the real part of the higher orders, and (d) u_1 with the imaginary part of the higher orders.

Figure 3.3(a) shows the first phase perturbation mode, v_1. Figure 3.3(b) shows the singular values, w. Figure 3.3(c) shows the first diffraction order amplitude variation

mode, u_1, where the central five orders are the amplitude and the higher orders are the real part. Figure 3.3(d) shows the same diffraction order amplitude variation mode with the imaginary part of the higher orders. The central orders are weighted higher than the other orders and therefore will vary more.

Inversion

Before SVD can be performed on the sensitivity matrix to construct the pseudoinverse, the sensitivity matrix must first be constructed. As equation (3.8) hints, the sensitivity matrix is composed of amplitude change to phase change ratios, $dA_{kl}/d\Phi_{mn}$. The matrix is constructed by actually perturbing the phase at each CGH phasel and recording the amplitude change. The process is expedited by utilizing a look-up table of analytically calculated complex far-field amplitude changes resulting from the perturbations at each phasel. Weighting factors for the desired amplitude pattern, f_{mn}, are also encoded in the sensitivity matrix which would then look like Figure 3.4. The number of unknowns, N, is the number of CGH phases and is set once the unit cell size is chosen. The number of measurements, M, is the number of diffraction order amplitudes

$$
\mathcal{H} = \begin{pmatrix} \dfrac{dA_{11}}{d\Phi_{11}} f_{11} & \cdots & \cdots & \dfrac{dA_{1N}}{d\Phi_{1N}} f_{1N} \\ \cdots & & & \cdots \\ \cdots & & & \cdots \\ \dfrac{dA_{M1}}{d\Phi_{M1}} f_{M1} & \cdots & \cdots & \dfrac{dA_{MN}}{d\Phi_{MN}} f_{MN} \end{pmatrix}
$$

Figure 3.4: Representation of the contents of the CGH sensitivity matrix.

that are calculated in the far-field, and is set by however many orders are desired. It then becomes apparent that this construction lends itself very easily to a polychromatic design - amplitudes at several different wavelengths are determined and the results are placed in the sensitivity matrix. Designs that have been fabricated typically use five optimization wavelengths within one octave of a particular band (i.e. from 400 nm to 800 nm).

The algorithm begins from a random phase CGH and the resulting far-field amplitudes are calculated. The sensitivity matrix is then constructed. Taking the difference between the current amplitudes and the desired amplitudes forms an error vector, $\Delta \mathbf{A}_e$. The least squares solution to the inversion problem is found by applying the pseudoinverse to the error vector:

$$\delta \Phi = -\mathcal{H}^+ \Delta \mathbf{A}_e = -\mathbf{V} \tilde{\mathbf{W}}^{-1} \mathbf{U}^t \Delta \mathbf{A}_e \tag{3.20}$$

The contribution of each phase perturbation mode is proportional to the inverse of its corresponding singular value (or zero for the modes beyond the rank). For modes with the smallest singular values, a large phase change is needed to produce a small amplitude change, and these are the modes that introduce nonlinearities. This can be reduced by thresholding the singular values. As mentioned in the linearity section, a more clever way of reducing the nonlinearity that was developed by Volin is simply to constrain the real and imaginary parts of the low irradiance (i.e., undesired) orders separately. This in effect relates the phases to the complex fields, not their amplitudes, and avoids the nonlinearities of taking the absolute values of the electric fields. No thresholding is necessary in this case, though there are still nonlinearities due to relating the phase, rather than the exponential of the phase, to the complex field. The presence of nonlinearities in

the algorithm prohibits a single operation of inversion from arriving at an optimized solution. The optimized solution is arrived at by iterating the above procedure, where each iteration depends on a recalculation of the sensitivity matrix. In the computer this is fast, but the iteration time becomes a factor when the algorithm is implemented in feedback with the OPA.

SVD design examples

It is useful to go through a few design examples. The first will be the 2D CGH design for visible wavelengths for the CGH of Figure 3.1. The second example covers the design of a 1D CGH, which is of interest since the OPA is inherently 1D. The 1D design is also compared with an etched 1D CGH that was fabricated to model the OPA's dimensionality and pitch.

2D CGH design

The first step is to specify the design goal. One of the typical designs is to have increasing efficiency with each order's radial distance from the center. This is illustrated in Figure 3.5(a) and is termed the *uniform irradiance* design. The purpose of this design is to counteract the loss of energy that some FPA pixels see in the higher orders due to dispersion and thus make better use of the dynamic range of the camera. The weights

1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1	1	1	1
1	1	1	1	1	1.5	3.5	3.5	3.5	3.5	3.5	1.5	1	1	1	1
1	1	1	1	1	1.5	3.5	3.5	3.5	3.5	3.5	1.5	1	1	1	1
1	1	1	1	1	1.5	3.5	3.5	3.5	3.5	3.5	1.5	1	1	1	1
1	1	1	1	1	1.5	3.5	3.5	3.5	3.5	3.5	1.5	1	1	1	1
1	1	1	1	1	1.5	3.5	3.5	3.5	3.5	3.5	1.5	1	1	1	1
1	1	1	1	1	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1	1	1	1
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

(a)   (b)

Figure 3.5: (a) The goal amplitude pattern and (b) the weights applied to the orders in the algorithm.

that are applied to the orders in the algorithm are shown in Figure 3.5(b). There is a balance to be struck between the total efficiency in the desired pattern and the broadband match to the goal pattern when choosing the weights. Also, the efficiencies of all the desired orders are designed to be spectrally flat. Figure 3.6 shows the ideal spectral efficiencies with the zero order in red, the first orders in green, the second orders in blue, and the higher undesired orders (which will show up in the spectral efficiency plots of the actual designs) will be shown in black. There are only 6 efficiency lines for the 25 orders because of redundancy in some of the first and second orders. For example, the four corner orders are intended to have the same efficiency of just over 6.5%.
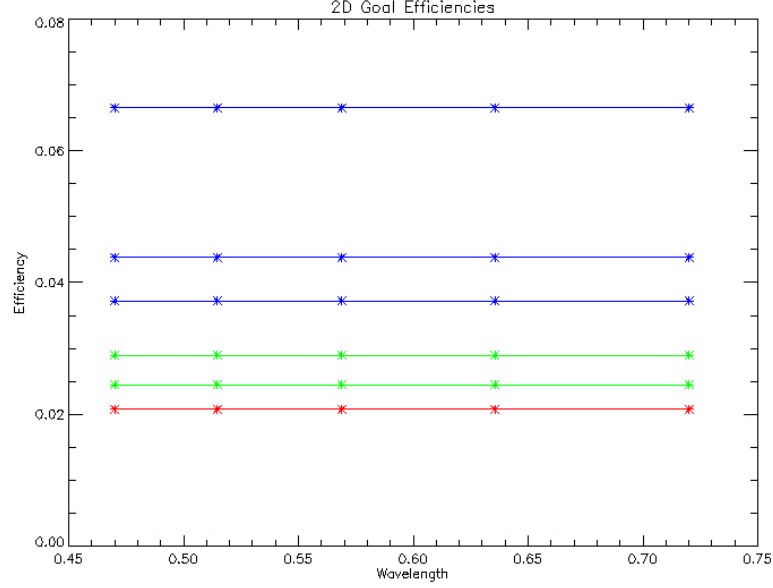
Figure 3.6: The goal efficiencies of the orders for the uniform irradiance design. For the 25 desired orders shown in Figure 3.5(a), there are six different desired efficiencies which increase with each order's radial distance from the center. The asterisks denote the design wavelengths.

There are also two numbers that are used to characterize the simulated performance of a particular design. These measure the match of the resulting diffraction orders to the goal pattern at the design wavelengths using a root-sum-square (rss), as well as the average of the efficiency of the desired orders over wavelength $\overline{\eta}_\lambda$. The root-sum-square is given by:

$$rss = \left(\sum_{k=0}^{K-1} \left(\frac{\mathbf{A}_k \mathbf{G}_k}{a\mathbf{F}_k} - \frac{\mathbf{G}_k}{\mathbf{F}_k} \right)^2 \right)^{\frac{1}{2}} \qquad a = \frac{\sum_{k=0}^{K-1} \frac{\mathbf{A}_k \mathbf{G}_k}{\mathbf{F}_k^2}}{\sum_{k=0}^{K-1} \frac{\mathbf{G}_k^2}{\mathbf{F}_k^2}} \qquad (3.21)$$

The diffraction order amplitudes of the desired orders and the real and imaginary parts of the far-field undesired orders at all the design wavelengths are organized into the vector **A**. The *goal* amplitudes of the desired orders and the real and imaginary parts of the far-field undesired orders at all the design wavelengths are arranged into the vector **G**. The design weights at the diffraction order locations at all the design wavelengths are organized into the vector **F**. The number *a* provides a scaling factor between the goal and calculated diffraction orders. The summation is over K points, which is greater than the number of the diffraction orders at all the wavelengths since the real and imaginary parts of the desired orders are separated. The smaller the rss, the more closely the design matches the goal. Designs that match the goal pattern well tradeoff total spectral efficiency, which is the second number:

$$\overline{\eta}_\lambda = \frac{1}{L}\sum_{l=0}^{L-1}\sum_{n=0}^{N'-1}\eta_{ln} \tag{3.22}$$

where L is the number of design wavelengths and N' is the number of desired orders. The efficiencies are calculated with respect to the light *leaving* the disperser and are not *absolute* efficiencies.

The first example design is the 2D CGH of Figure 3.1. The goal parameters are as stated above and a CGH cell size of q = 16 (16 x 16 phasels) is chosen. Figure 3.7 shows the spectral efficiency plots, and below the plot are images of the amplitudes of the orders at the design wavelengths (note that no effort has been taken to show the spatial expansion of the orders with wavelength due to dispersion). The rss and $\overline{\eta}_\lambda$ are listed along with the figures. The design is fairly typical of a 2D design with nearly 74% of the
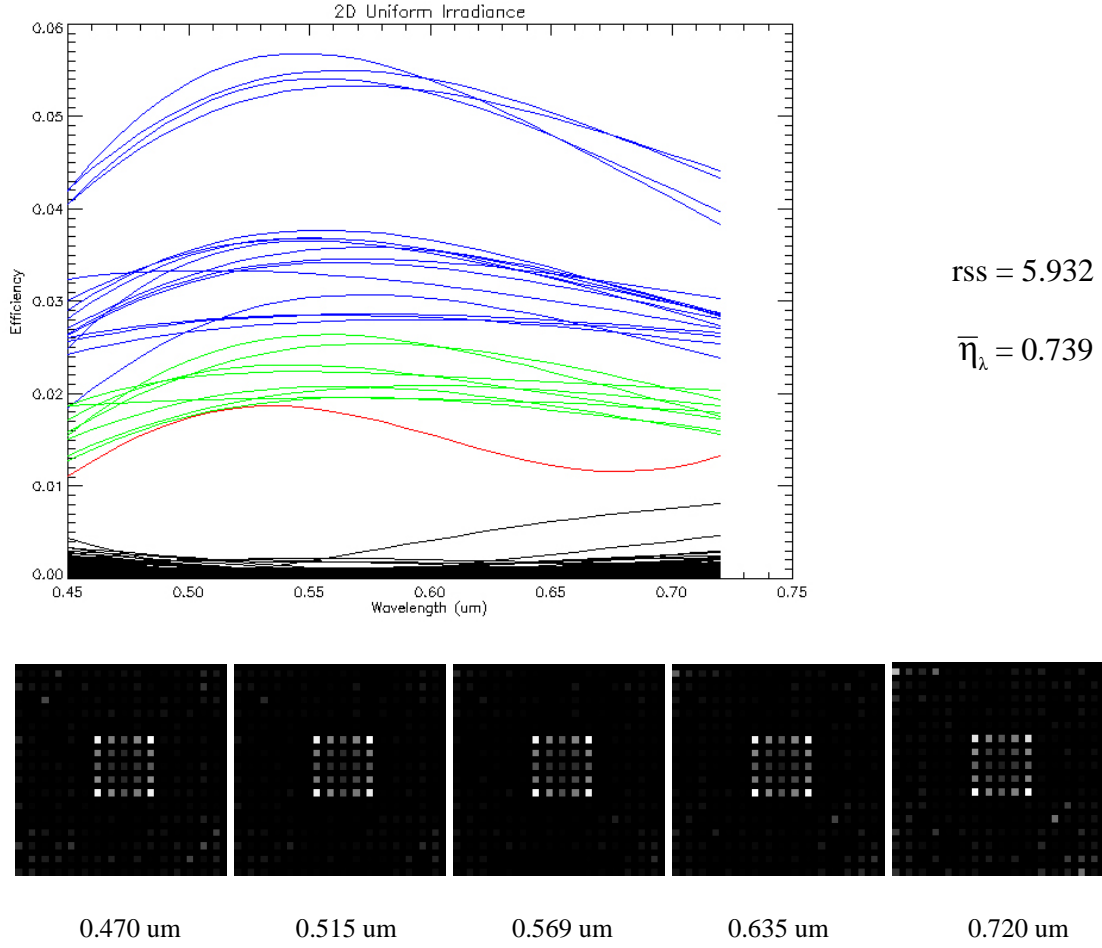
Figure 3.7: Plot of the spectral efficiencies of the diffraction orders for the 2D uniform irradiance design, with images of irradiances at the design wavelengths shown below the plot.

exiting light coupled into the desired pattern. The grouping of the orders at different efficiency levels is apparent in the plot, especially for the four corner orders. The rss measure is a useful comparison between results for different numbers of iterations or initial conditions, but is not useful for comparisons between different CGHs. For example, the rss between this design and one with the same goal but with 8 x 8 phasels (q = 8) cannot be compared.

<u>1D CGH design</u>

The next design is the 1D version with the same goal. The goal spectral efficiencies are shown in Figure 3.8(b). There are now only 5 desired orders and the goal efficiencies of the first and second orders are redundant. Again, the CGH cell size is chosen to be q = 16 (note that the phase is now a 16 x 1 vector for the 1D design). The results of the 1D uniform irradiance design are shown in Figure 3.9. In this case over 80% of the exiting light is coupled to the desired orders.
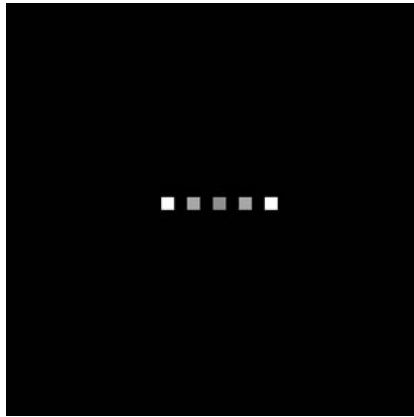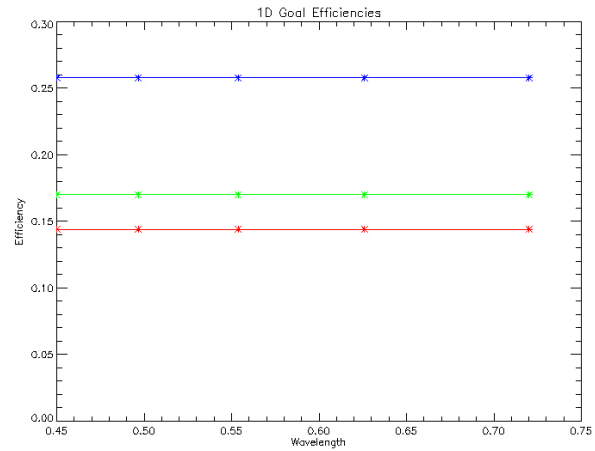
(a)   (b)

Figure 3.8: (a) The 1D goal amplitude pattern and (b) the spectral efficiencies of the diffraction orders for the 1D uniform irradiance design.
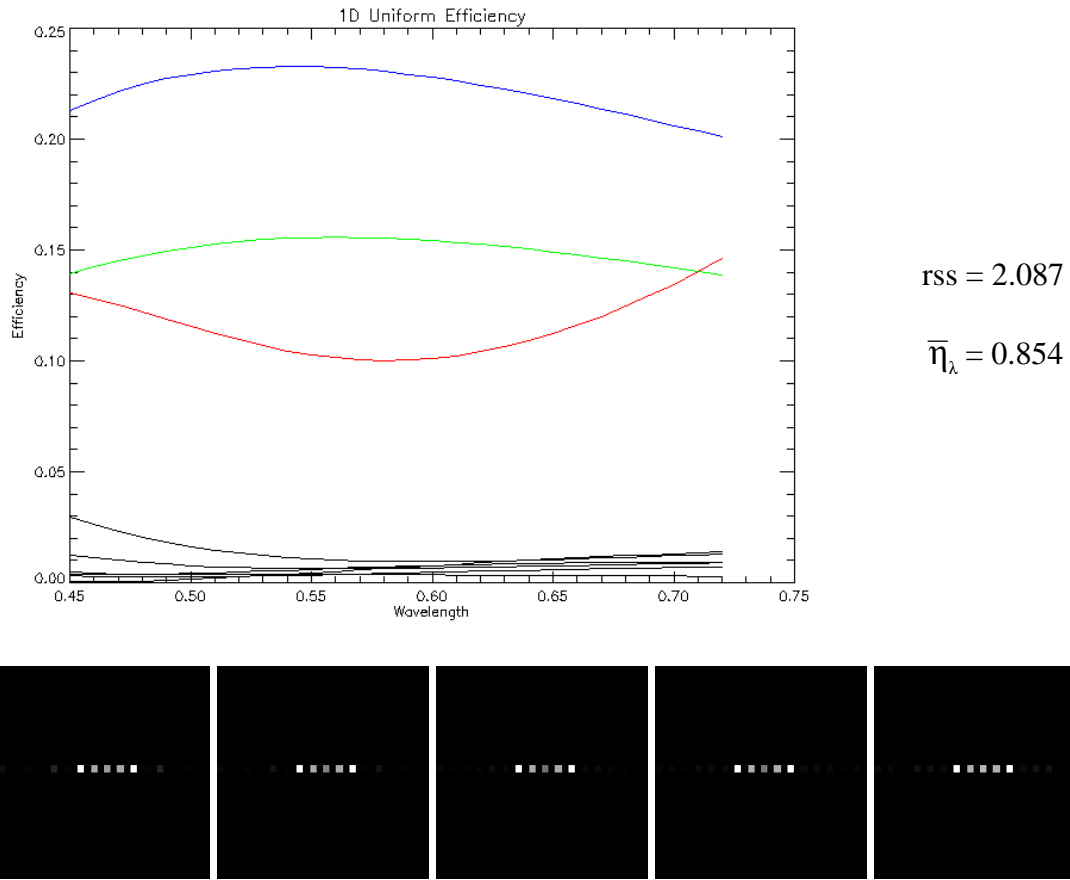
Figure 3.9: The spectral efficiencies of the diffraction orders for the 1D uniform irradiance design.

$$rss = 2.087$$

$$\overline{\eta}_{\lambda} = 0.854$$

<u>1D etched CGH measurement</u>

The 1D uniform irradiance design was used to make a pair of 1D etched CGHs. The purpose of these CGHs is to model the physical parameters of the OPA. The etched CGHs are identical and have 16 CGH pixels with a 2.5 um pitch. The CGHs were etched into PMMA plastic on a glass substrate and have a 1" diameter clear aperture.

<div align="center">(a)   (b)</div>
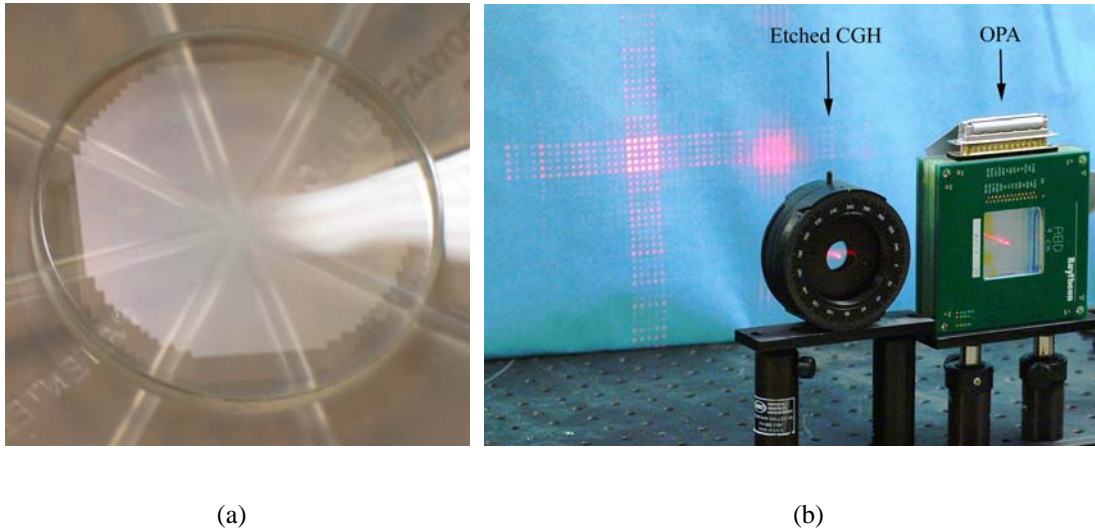
Figure 3.10: Photograph of the etched CGH on the left. On the right, both 1D etched CGHs are mounted in rotary mounts and are crossed to effect a rectangular diffraction pattern and are shown next to the OPA.

Figure 3.10(a) shows a photo of one of the etched CGH dispersers. Figure 3.10(b) shows both etched CGH dispersers in their rotary mounts and crossed to effect a rectangular grid of diffraction orders. The OPA is shown alongside the etched dispersers. The procedure for determining the efficiencies is very similar to the procedure for CTIS calibration and the experimental set-up is shown in Figure 2.3, only here a single etched CGH replaces the polarizer and OPA in the CTIS. First, a set of calibration images is taken covering the spectral range over which the diffraction efficiencies of the orders is desired. The images are processed by a dark subtraction and thresholding to eliminate fixed pattern noise. The thresholding becomes an issue for images with very low signal, but is not a problem here as the signal in the diffracted orders is nearly two orders of magnitude larger than the threshold value. The locations of the orders in the images are determined and the total light that is coupled to each order is calculated as the sum of a

small pixel area (typically 50 x 50) surrounding that order. The absolute efficiencies are obtained by dividing each spectral image by the incident light for each wavelength, which is acquired by simply repeating the acquisition with the CGH removed. These images containing the incident light are also subjected to a dark subtraction and threshold.

Figure 3.11 shows the spectral efficiencies of the etched CGH. The blue lines are the second orders, the green lines are the first orders, and the red line is the zero order. A quantitative comparison of the etched CGH to the simulation of Figure 3.9 is difficult because the *absolute* efficiencies of the etched CGH are being compared to the efficiencies of the design orders for the light *exiting* the CGH, the difference being the reflection, absorption, and scattering of the etched CGH. The diffraction order
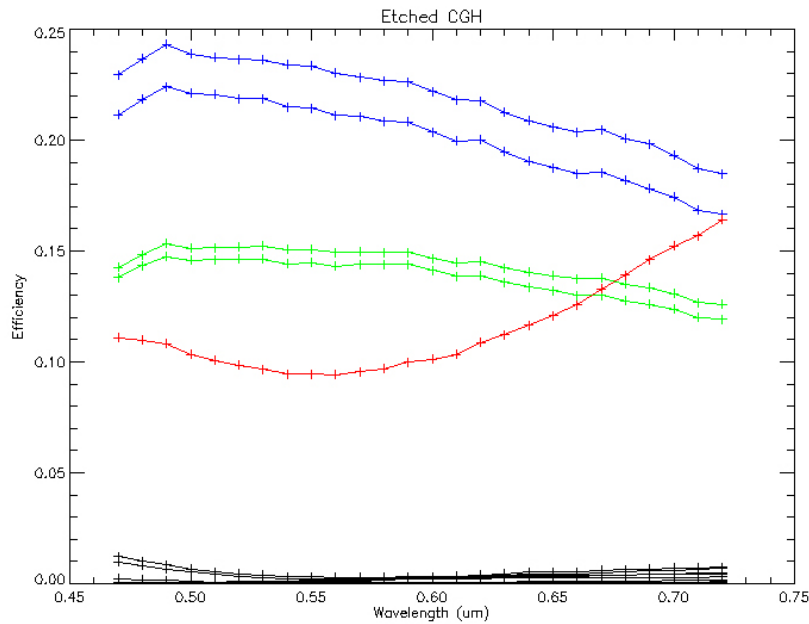
$$\overline{\eta}_{\lambda} = 0.815$$

Figure 3.11: Spectral efficiencies of the 1D etched CGH disperser.

efficiencies are not symmetric due to slight fabrication errors. Also, with a 2.5 um phasel pitch (d = 5λ), the CGH is on the edge of the scalar diffraction model's validity. Historically, the evaluation of the measured absolute efficiencies as shown in Figure 3.11 is only qualitatively compared with the simulation of the efficiencies of the light *leaving* the CGH as shown in Figure 3.9. The most noticeable differences between the two are the asymmetry of the spectral efficiency curves of the matching orders in Figure 3.11 and the slight deviation from the desired performance at high wavelengths (less light is coupled to the higher orders as compared with the simulation). The total spectral efficiency average of the measured absolute efficiencies in the desired orders is only 4% less than the simulation. Qualitatively speaking, this is a very good CTIS CGH with performance that closely matches the idealized simulation.

SVD design convergence

A useful experiment is to constrain the SVD CGH design algorithm to operate as it would in feedback with the OPA. This means that the real and imaginary parts of the dim orders cannot be constrained separately, and the goal of this simulation experiment is to gain some insight into the convergence of the algorithm. The convergence of the algorithm will depend on the behavior of the phase correction vector, equation (3.20), and the rss, equation (3.21), as the algorithm iterates. If the phase correction vector maintains large values, the algorithm is not properly constrained and will not converge. On the other hand, if the phase correction vector approaches zero as the algorithm iterates, it has stabilized in at least a local minimum in the solution space. The rss is a measure of the

solution's match to the goal and indicates whether the algorithm has settled on a *reasonable* solution and is really doing what it is supposed to be doing.

The phase correction vector provides information about the *convergence* of the algorithm and the rss provides information about the algorithm's *performance*. Four different versions of the algorithm are compared, each of which varies only its constraints on the phase correction vector. The first version is the baseline algorithm where the real and imaginary parts of the low irradiance orders are constrained separately with no thresholding. This corresponds to the 1D CGH design example of Figure 3.9. The second version uses the amplitudes of all the orders with no thresholding. The third version uses the amplitudes of all the orders and thresholds the smallest singular value by setting it to zero. The fourth version uses the amplitudes of all the orders and a threshold value is chosen such that two or three of the smallest singular values are set to zero. Remember that the phase correction vector is proportional to the inverse of the singular values, so by thresholding small singular values the magnitude of the phase correction vector is limited. The design goal in all four versions is identical to the 1D CGH design example with 16 phasels and has the same efficiency goal pattern as shown by the spectral efficiency plots of Figure 3.8. All four versions are plotted in Figures 3.12(a)-(d). The red line is the first version (baseline), the black line is the second version (all amplitudes), the green line the third version (small threshold), and the black line is the fourth version (larger threshold). Figure 3.12(a) plots the average of the modulus of the phase correction vector (in radians) as a function of iteration number. This is calculated according to equation (3.23):

$$\left|\delta\overline{\Phi}\right| = \frac{1}{q}\sum_{m=0}^{q-1}\left|\delta\Phi_m\right| \tag{3.23}$$

Notice that Figure 3.12(a) is a semi-log plot. This number indicates the magnitude of the

phase correction vector change from one iteration to the next. If it stays high, as does the

black line of version two (amplitudes with no threshold), the algorithm is not well

(a) $\left|\delta\overline{\Phi}\right|$ (b) $\sigma_{\delta\Phi}$

(c) - rss   (d) - $\overline{\eta}_\lambda$
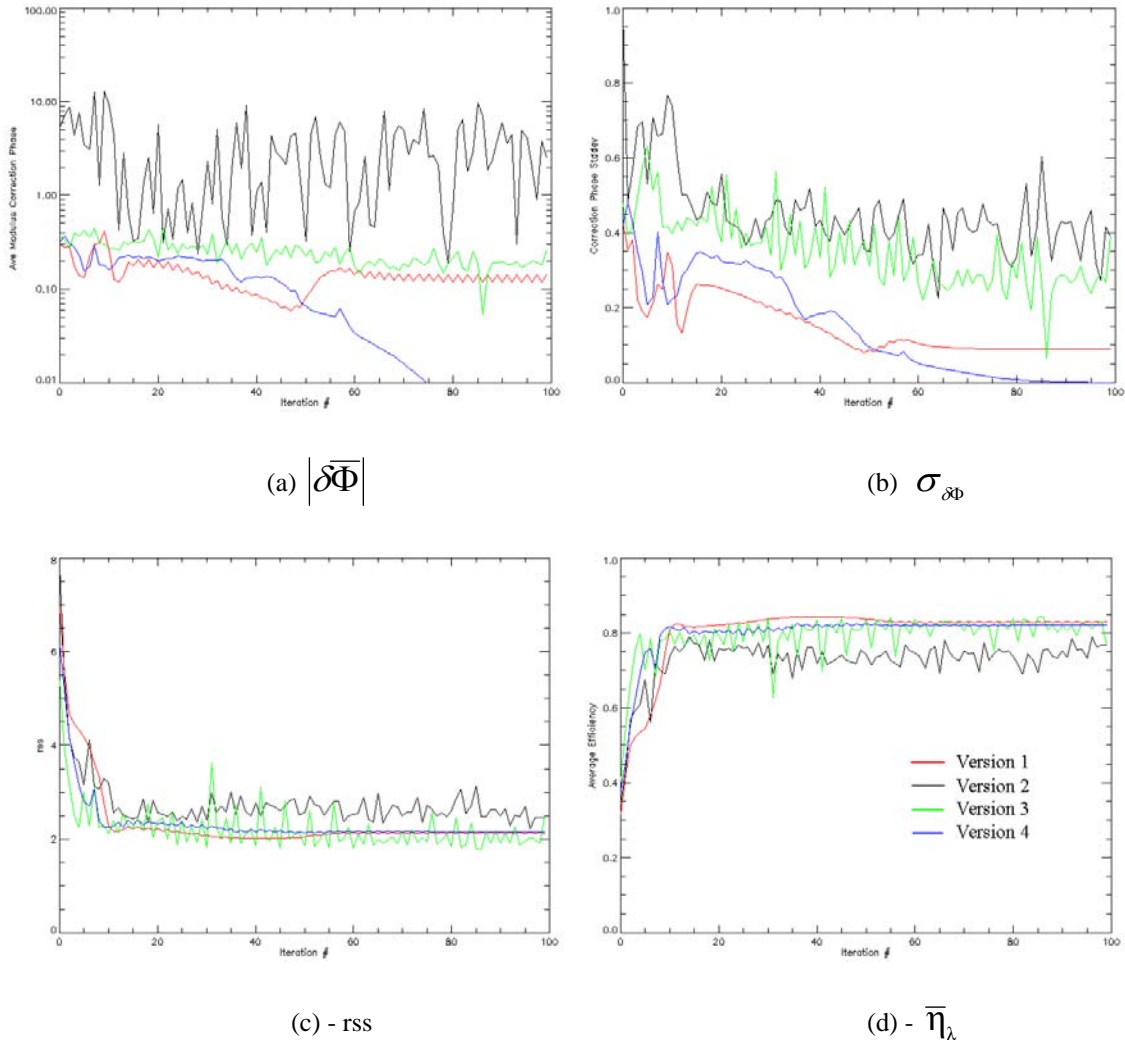
Figure 3.12: (a) Average modulus of the phase correction vectors, (b) standard deviation of the phase

correction vectors, (c) rss of the four designs, and (d) $\overline{\eta}_\lambda$ of the four designs.

constrained. It turns out that in this case version two still converged. Figure 3.12(b)

plots the standard deviation of the phase correction vector. This number indicates the

uniformity of the phase correction vector change, but does not give significant

information about algorithm convergence. Notice that in all the plots, the first and fourth

versions are very smooth while the second and third versions are 'noisy.' This is due to

nonlinearities in these versions. The first version reduces nonlinearities by separately

constraining the real and imaginary parts of the dim orders while the fourth version

reduces the nonlinearities by thresholding the small singular values. The second and

third versions are susceptible to nonlinearities at each iteration due to the small singular

values that are retained. Figure 3.12(c) plots the rss of each version. This plot indicates

that the nonlinearities do not necessarily prevent the algorithm from finding an optimum

solution, though too much non-linearity error does have ill effects as can be seen from the

plot for version two (the black line). Figure 3.12(d) plots the average efficiency, $\overline{\eta}_\lambda$, for

each version. The results here are similar to the rss – versions one and four give both

high efficiency and converge smoothly, version two gives inferior (yet still 'good')

results, and version three gives high efficiency but does not converge smoothly.

The main conclusion of this simulation is that CGH design results similar to the

baseline algorithm (version one) are achievable without access to the complex electric

fields. CGH performance similar to etched CGHs (which represent the current state-of-

the-art CTIS disperser) is theoretically attainable with the OPA. The experiment

confirms that it will be necessary to threshold small singular values when the algorithm is

used in feedback with the OPA to obtain a smoother convergence. Smooth convergence

is important because there are a myriad of effects when implementing the design with the OPA that can throw the algorithm into wild behavior such as low signal, FPA noise, stray light, source stability, and the non-linear phase response of the OPA and its crosstalk. The limit of the thresholding was not explored. It is expected that performance falls off with increased thresholding due to over-constraining the phase correction vector – i.e. the necessary information to perform the inversion will have been eliminated.

CHAPTER 4 – OPTICAL PHASED ARRAY

OPA architecture

The tunable CTIS utilizes a transmission liquid crystal spatial light modulator (LC SLM) as its tunable disperser. The LC SLM is borrowed from Raytheon Network Centric Systems (RNCS). Raytheon created the device for beam steering applications, and thus they refer to the LC SLM as an Optical Phased Array (OPA) in direct analogy with traditional microwave phased arrays. This particular device was fabricated in the early 1990s and is used as a portable demonstration device. Figure 4.1 shows the OPA with its 37 pin connector plugged into its electronics base, which also is used to mount a polarized HeNe laser.
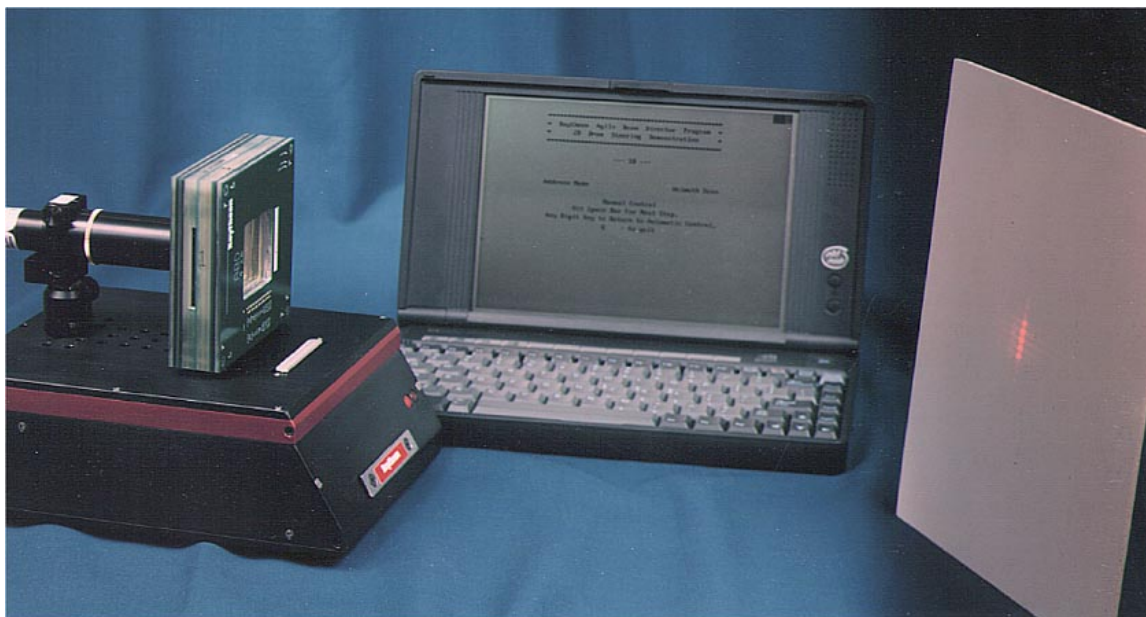
Figure 4.1: The Raytheon OPA. (Courtesy of Raytheon Network Centric Systems)

A computer communicates with the electronics base via 10 base T raw ethernet packets. The drive and control electronics were custom built for an 80486 based computer with an EISA bus. The control has been adapted for use with the laptop seen in Figure 4.1 and the communications conform to an Ethane protocol. It is uncertain what type of protocol Ethane is, and EISA (along with its drivers that the control program seems to use) is now obsolete. Consequently, a good deal of effort was spent in developing communications with the OPA using current software. Appendix A covers the details of the custom communications software that was developed for this project.

For beam steering applications, Raytheon uses two OPAs. One is 'coarse' and one is 'fine' in reference to coarse and fine beam steering. Only the coarse OPA is used for this project and is shown in Figure 4.1. The coarse OPA has an electrode width of 1 um. The clear aperture of the OPA is 4 cm square. Given this large aperture and such a fine pitch there is a very large number of individual electrodes. It is difficult to both fabricate and address such a large 2D pattern of electrode pixels, so Raytheon fabricated the device as two orthogonally crossed 1D LC cells that are independent of each other. To further ease the addressing of the OPA electrodes, each cell has 256 independently addressable electrodes that form a unit period that is repeated 84 times across the 4 cm aperture.

The OPA works on the same principal as the etched CGH gratings – a phase profile pattern is realized by creating a spatial optical path distance pattern, OPD = n*d. The etched CGH changes d, while the OPA changes n. The OPA uses nematic E7 liquid crystal. This liquid crystal material contains anisometric 'cigar' shaped molecules.

Normal fluids are isotropic in orientation so that even though their molecules may be anisometric in shape, the macroscopic optical, magnetic, electrical, etc., properties are isotropic. Nematic liquid crystals exhibit orientational order as in a solid yet still flow like a liquid. The average molecular orientation is termed the director. Smectic liquid crystals exhibit even more order and the flow is constrained to separate layers. Figure 4.2 shows a schematic of the three example states. The orientational order of the anisotropic molecules of the nematic liquid crystal lead to anisotropic optical properties such as birefringence, which is the mechanism used to change the index of the liquid crystal material to achieve phase shifting with the OPA.

| Isotropic | Nematic | Smectic |
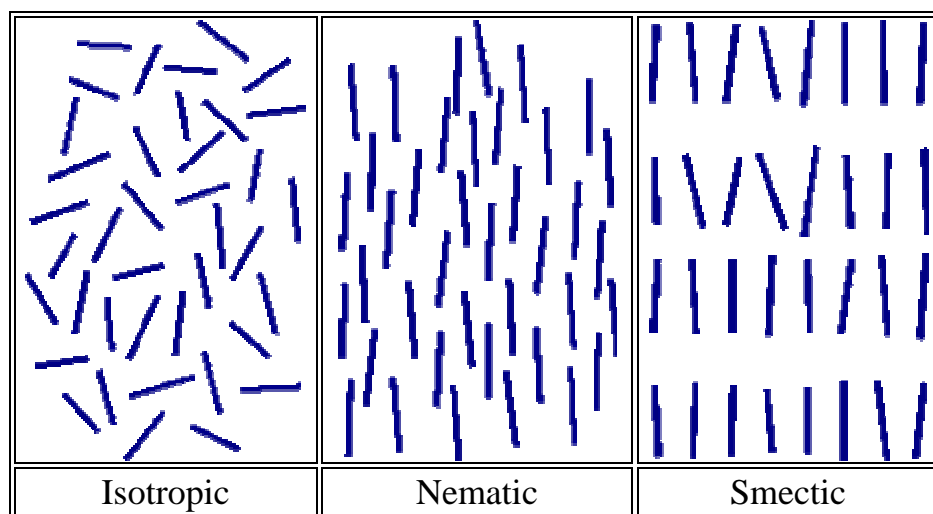
Figure 4.2: Isotropic, nematic, and smectic molecular states. The nematic, with its anisotropic optical properties, is the type used in the OPA.[19]
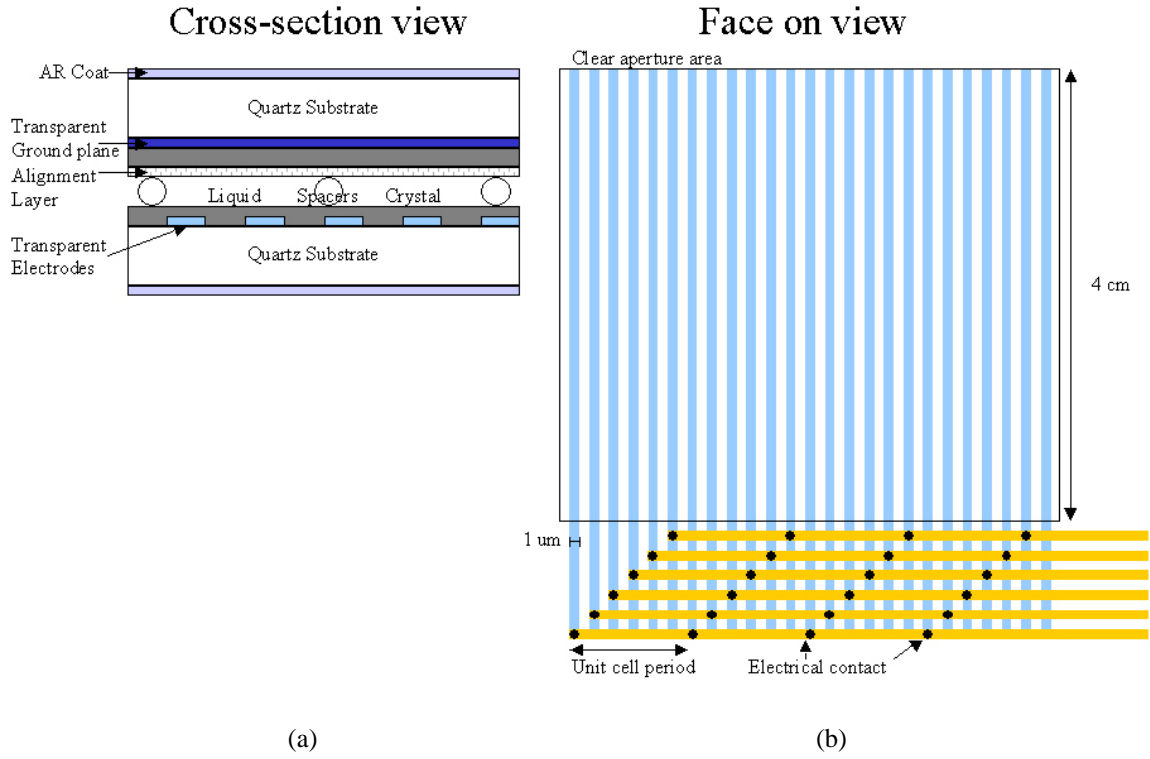
Figure 4.3: (a) Cross-sectional and (b) face on schematic views of the OPA.[7]
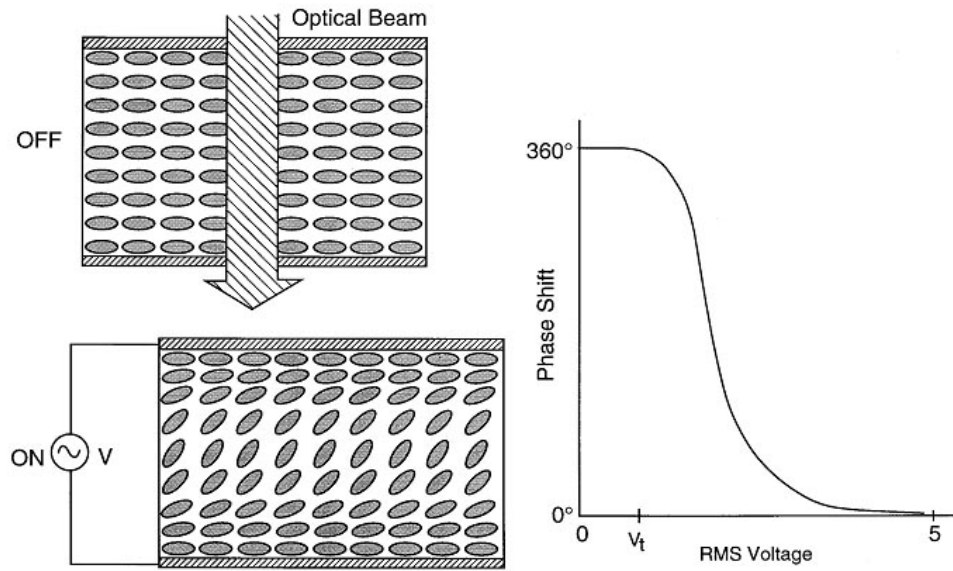
Figure 4.4: Schematic of the 'on' and 'off' positions of a single OPA electrode and the non-linear phase response with applied voltage.[7]

The liquid crystal cell is formed by sandwiching a thin layer of liquid crystal between two substrates, which is schematically represented in Figure 4.3(a). The outside of one of the substrates is patterned with the transparent electrodes, and a transparent ground plate is laid across the entire area of the outside of the other substrate. The insides of the substrates are scored with an alignment layer that gives a preferential orientation to the molecules adjacent to its surface. Mechanical forces within the material align the rest of the molecules in the volume. The component of an incident optical field parallel to the long axis of the liquid crystal molecules will then see the extraordinary index of refraction, n_e, of the liquid crystal. This is illustrated in the 'off' position of Figure 4.4.

Applying a voltage to the electrodes of the OPA causes the liquid crystal molecules to align with the electric field, which alters the index towards that of the ordinary index, n_o, for that same component of polarization. The voltage at which the molecules begin to align with the field is the threshold voltage, V_t, and the voltage at which the molecules are completely aligned with the field is the saturation voltage, V_s. The dynamic phase range that the OPA can achieve is given by a modified form of equation (3.1):

$$\Delta\Phi_\lambda = \frac{2\pi(n_e - n_o)d}{\lambda} \tag{4.1}$$

The birefringence, $\Delta = n_e - n_o$, is 0.2 for nematic E7. To obtain a phase shift of one wave, d/λ needs to be 5. The original beam steering application for the OPA was at a wavelength of 1.06 um, and thus it was made with a thickness near 5 um. The molecular

orientation is responsive to the rms value of the applied voltage, which is a sinusoid in

the 1 to 10 V range at 1 to 10 kHz.

Response time

One of the major features of liquid crystal materials is switching speed. How fast

the LC can change from one molecular orientation to another depends on the liquid

crystal's viscosity, γ, its effective relaxation constant, κ, the operating temperature, the

applied voltage, and the thickness of the cell. The Leslie-Ericksen equation[20] is used with

some simplifying assumptions to define the decay time and rise time of the liquid crystal

director as:[7]

$$\tau_d = \frac{t^2\gamma}{(\kappa\pi^2)}$$

$$\tau_r = \frac{\tau_d}{\left[(V/V_t)^2 - 1\right]}$$

(4.2)

These equations are for a uniform phase profile. Even so, the quadratic scaling of the

decay time with thickness, t, is still valid for more complex profiles. For most cases, the

decay time limits the liquid crystal switching speed. Increasing the operating temperature

decreases the viscosity, and thus decreases the switching speed up to the point where the

phase transition temperature to that of an isotropic liquid is reached. For E7, this

temperature is near 61^0 C. For the OPA, the switching speed is near 4 ms and will not

limit the speed of the CGH design algorithm. The speed limit for the algorithm is set by

the frame rate of the camera, which is 15 fps, or 66 ms.

Flyback

One of the major difficulties of using fine pitch liquid crystal spatial light modulators is inter-electrode crosstalk, which is often referred to as 'flyback' The term flyback originates from beam steering applications in which the ideal phase profile is a ramp that uses 2π phase resets to keep the needed phase dynamic range to reasonable levels. The result is a sawtooth phase profile. Flyback is the inability of the device to do a 2π phase reset between adjacent electrodes, causing a reduction in the efficiency of the steered beam. Figure 4.5 depicts a simple phase versus position plot for a single period of a phase sawtooth wave that illustrates flyback. Equation (4.3) shows the approximation of the efficiency of the desired order for a phase ramp that exhibits flyback.[7] The duty cycle of the sawtooth wave is Λ, and the flyback distance is Λ_F.

$$\eta = \left(1 - \frac{\Lambda_F}{\Lambda}\right)^2 \tag{4.3}$$

There are two root causes of the inter-electrode crosstalk. The first is the complicated fringing electric fields within the liquid crystal media that affect the orientation of the director in the spatial area of neighboring electrodes. The second is the mechanical stress on the LC molecules due to reoriented nearest neighbors. Figure 4.6 illustrates the extreme case of crosstalk with the desired phase between adjacent electrodes switching between the entire dynamic range.
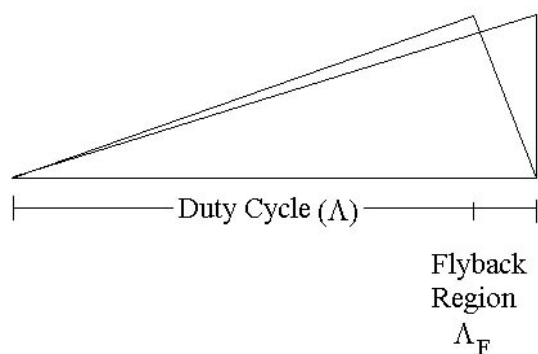
Figure 4.5: A single period of the sawtooth phase ramp illustrating flyback.[7]
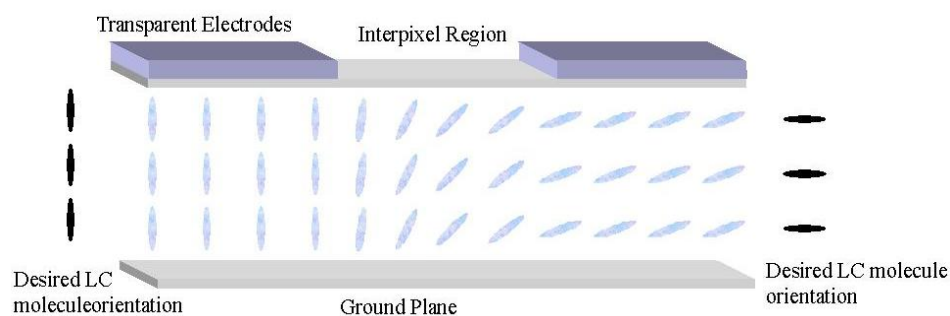
Figure 4.6: Illustration of flyback as a type of inter-electrode crosstalk between the molecules of adjacent electrodes due to complicated fringing electric fields.

A rule of thumb is that the lateral extent of the crosstalk is equal to the liquid crystal cell thickness. For the OPA, this means that the applied voltage at one electrode can affect the orientations of the molecules, and thus the phase of the incident light, up to two electrodes away. This is the primary reason that the desired phase profiles cannot be designed a priori in the computer and simply programmed to the OPA using the appropriate voltages. Flyback forces the use of some type of feedback algorithm where

the OPA electrode voltages are changed based on the *detected* irradiances by the CTIS camera. Another factor that prevents the direct programming of phases to the OPA is the previously mentioned non-linear phase response with applied voltage, which is the subject of the next section.

OPA phase response

The phase delay of the OPA is not a linear function of the applied voltage. Generally the response follows the curve shown in Figure 4.4. Obtaining the phase vs. voltage response of the OPA can be done by using a common path interferometer, and this was done using a WYKO 6000 laser based Fizeau. The OPA was placed in the collimated space between the WYKO and a reference flat that was adjusted to introduce a few waves of tilt parallel to the electrodes of the LC cell under test. The voltages of the OPA electrodes are all changed together, and the resulting straight-line fringes travel in the direction perpendicular to the electrodes. Whenever a fringe travels a full period, the voltage has caused a phase delay of 2π at a wavelength of 633 nm (the WYKO uses a
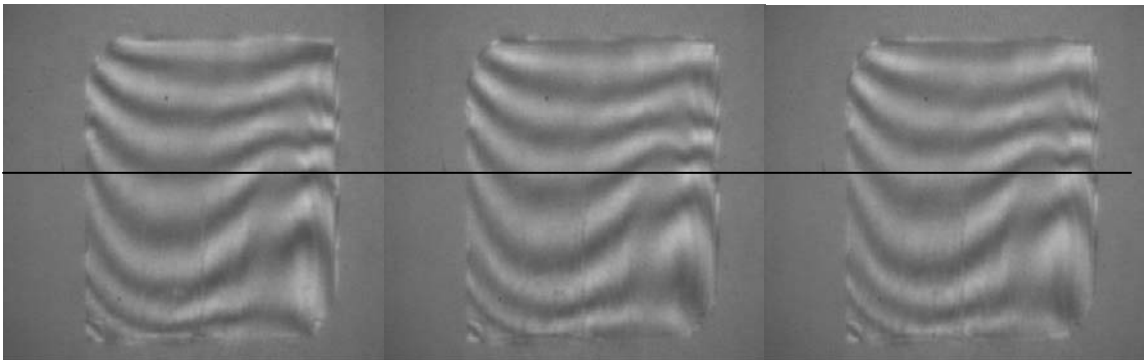
Figure 4.7: Fringe shifts in the WYKO experiment due to changing the voltage of the horizontal OPA electrodes.

HeNe laser for its illumination).   Figures 4.7(a-c) show the fringe shifts of the WYKO experiment.

Qualitatively, this test showed between two and three waves of dynamic range at this wavelength for the OPA. However, the quantitative results were inconclusive due to unacceptable fringe drift. The WYKO is not set up to automate this process and it takes a human controller close to 40 minutes to get a reasonable data set to infer the phase vs. voltage response. The data acquisition needs to take place within a minute or two to avoid the fringe drift. The common path interferometer is *extremely* sensitive to air currents and temperature changes. An easier approach is to calibrate the OPA as a binary phase grating.

<u>Binary phase grating</u>

A binary phase grating is completely transmitting and has only two phase levels. Thus it is easy to analytically calculate the far-field amplitudes of the *ideal* binary phase grating. The ideal binary phase grating has a square wave phase profile in 1D, as shown
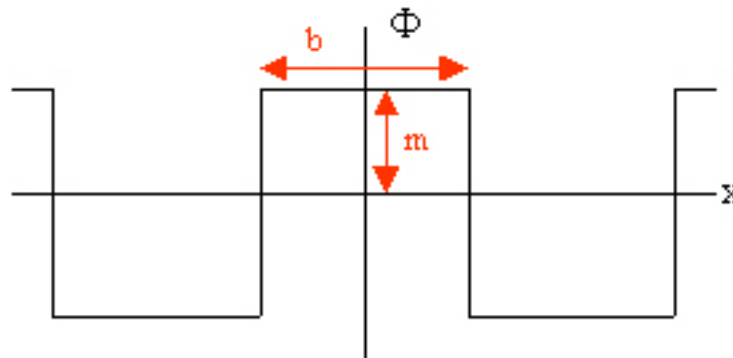
Figure 4.8: The ideal 1D binary phase profile.

in Figure 4.8. The OPA cannot perform as the ideal binary phase grating (nothing can

because of the discontinuity at the transition between the two phases) due to flyback.

However, when the maximum period available with the OPA is used (256 electrodes), the

ideal binary phase grating serves as a good approximation. The transmission of the ideal

binary phase grating is:

$$t(x) = \left(e^{-i2\pi m} rect\left(\frac{x}{b}\right) + e^{+i2\pi m} rect\left(\frac{x-h}{b}\right) \right) * \frac{1}{2h} comb\left(\frac{x}{2h}\right) \tag{4.4}$$

where m = ½ for a full 2π phase delay, * denotes the 1D convolution, and the comb

function is defined by Gaskill[21] as:

$$comb\left(\frac{x}{h}\right) = |h| \sum_{n=-\infty}^{\infty} \delta(x - nh) \tag{4.5}$$

Given a normally incident, unit amplitude plane wave, the complex electric field in the

Fraunhofer region is the Fourier transform of equation (4.4).

$$E(\xi) \propto \left[\left(e^{-i2\pi m} + e^{i2\pi m} e^{-i2\pi h\xi} \right) bsinc(b\xi) \right] comb(2h\xi) \tag{4.6}$$

The diffraction order locations are at $\xi = n/(2h)$, where n is an integer ranging from

positive to negative infinity. Equation (4.6) then reduces to:

$$E_n \propto \left(e^{-i2\pi m} + e^{i2\pi m} e^{-i\pi n} \right) \frac{1}{2} sinc(n) \tag{4.7}$$

Note that the weight of the delta functions in the comb is 1/(2h) from the definition. The

diffraction efficiency of the orders is given by the squared modulus of equation (4.7):

$$\delta_n = \cos^2\left(2\pi m - \frac{\pi}{2} n \right) sinc^2\left(\frac{n}{2}\right) \tag{4.8}$$

Equation (4.8) gives the diffraction order efficiencies of the ideal binary phase

grating as a function of the phase difference (m) between the two levels. The zero order

(n = 0) efficiency is a cosine-squared function of the phase difference. All the higher

even orders do not exist as they fall on the zeros of the sinc function. The odd diffraction

orders are symmetrical and they follow the same cosine squared function as the zero

order except that they are shifted by π/2, or equivalently they follow a sine-squared

function rather than a cosine-squared function.
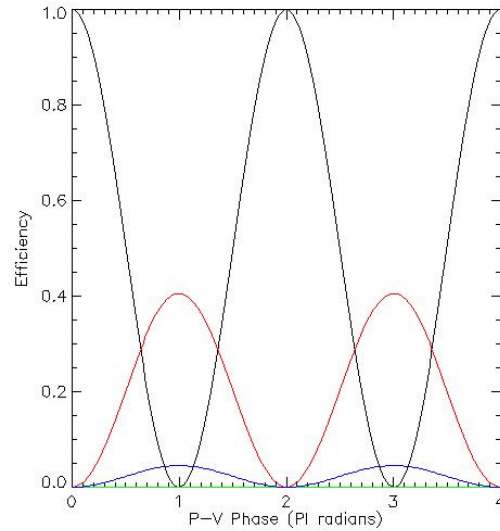
Figure 4.9: Efficiency as a function of phase difference for the ideal binary phase grating. The zero order is black, the first orders are red, the second orders are green, and the third orders are blue.

The amplitude of the cosine squared functions of the higher odd orders is

determined by the sinc-squared function, and nearly 82% of the light is coupled to the

first orders when m = ¼ (the peak-to-valley phase difference is π). At this phase

difference, the zero order vanishes. Figure 4.9 shows the efficiencies of the zero, first,

second, and third orders of the ideal binary phase grating as a continuous function of the

phase difference. The zero order is shown in black, the first orders are red, the second

orders (a constant zero is this case) are green, and the third orders are blue.

The binary phase experiment with the OPA is, once again, very similar to

calibration. The polarizer and OPA are placed in the CTIS as shown in Figure 2.3, and a

spatial/spectral point source is placed on axis in the field stop using the calibration

apparatus. The CTIS camera acquires the resulting diffraction orders. The voltages that

are applied to the first 128 adjacent electrodes of the OPA are held at a constant level

while the other 128 adjacent electrodes are changed uniformly to some other voltage.

Keep in mind that this pattern is repeated 84 times as there are 84 of these 256 electrode

unit cells in the construction of the OPA – this is the binary square wave. The irradiance

of the acquired diffraction orders is normalized by the irradiance when all the electrodes

are held to a constant voltage (i.e. $m = 0$). The resulting data set contains the efficiencies

of the orders. This is repeated 199 times while increasing the voltage through its entire

dynamic range. The data for the efficiency of the any of the orders can then be fit to

equation (4.8), and this then determines the phase delay of the OPA electrodes as a

function of applied voltage. The fit has been done with both the zero and first orders, and

the results presented here use the first order. The reason for this is to get a smoother 'cut

on' transition for the phase vs. voltage curve, as this is difficult to do using the zero order.

As mentioned previously, the OPA was constructed for use at 1.06 um

wavelength light, and thus it has an anti-reflection coating optimized for this wavelength.

It turns out that this coating causes the transmission of the OPA to fall off rapidly

between 500 nm and 600 nm. The ideal wavelength range is between 450 nm and 720

nm, however, this becomes impractical due to low transmission at the lower wavelengths.

Thus a wavelength band from 550 nm to 720 nm is considered instead. Figure 4.10(a)

shows the results of the binary phase experiment at $\lambda = 625$ nm, and Figure 4.10(b)

shows the resulting phase vs. voltage response using a fit to equation (4.8) with the

negative first order.

The results shown in Figure 4.10 plot the efficiency and phase versus 'voltage

number' rather than voltage with units of volts. The control program sends the voltage to

be applied to the OPA as a 10 bit number which the electronics base then translates into

an actual rms voltage amplitude. The voltage amplitude scales linearly with the

quantized number that controls it, and the plots shown in Figure 4.10 simply use

(a)   (b)
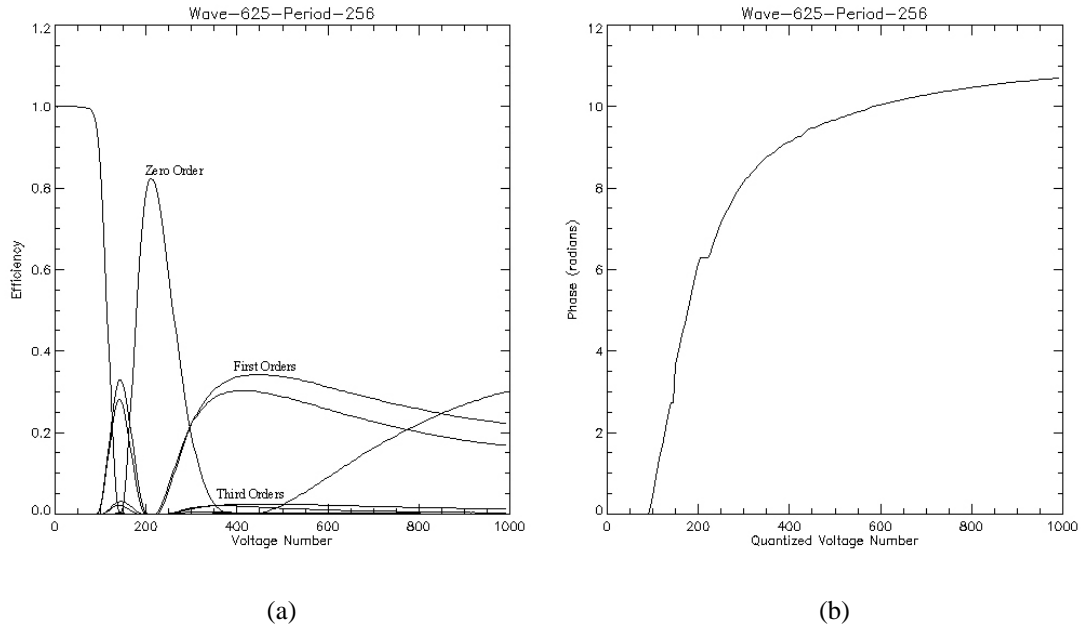
Figure 4.10: (a) Binary phase data for the first three orders and (b) the phase vs. voltage fit for the zero
order.

that quantized number for the abscissa. The experimental data set consists of 199 images taken at every 5th voltage number, thus the data set contains information from voltage number 0 – 990. The last six data points up to 1020 are not taken due to the particularities of the device stated in Appendix A.

It is useful to point out a few things about the data in Figure 4.10(a). Since there appears to be about 2 periods of the cosine-squared function of the zero order, the dynamic range at this wavelength appears to be about 4π. Notice that the zero order efficiency does not return to 1.0 (which is 100%) at 2π phase delay. This is due to a small amount of flyback for this period. The phase vs. voltage number for this wavelength is plotted in Figure 4.10(b). It is reasonable to assume that the results simply scale with wavelength, and the experiment was identically carried out at λ = 550, 575,

(a)   (b)

Figure 4.11: The phase vs. voltage number response (a) at λ = 550, 575, 600, 625, 650, 672, 700, and 720 nm (top to bottom, respectively) and (b) all scaled to the 625 nm wavelength.

600, 625, 650, 675, 700 and 720 nm. The data for each experiment was normalized for the 625 nm wavelength and plotted in Figure 4.11(b) for comparison. The jog in the data at 2π phase (near a voltage number of 200) is due to the fit of the data. An argument could be made that the scaling assumption is valid up to 2π phase delay, but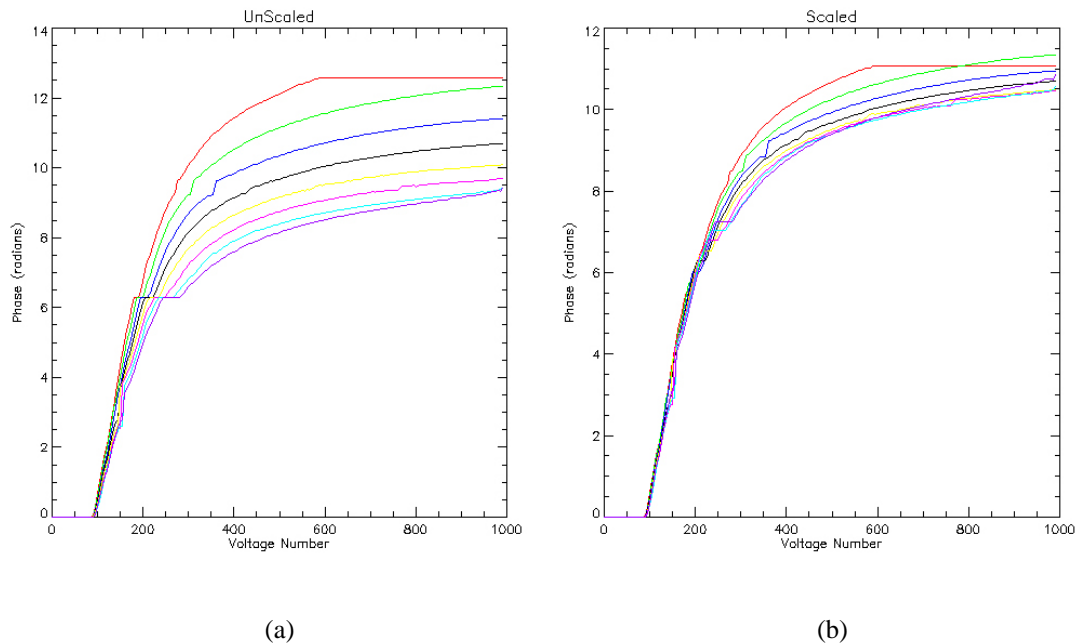 the results show that phase vs. voltage response does not satisfactorily scale with wavelength. The reason for this is found in the anti-reflection coating of the OPA. It is thought that there are etalon effects that take place within the AR coating at visible wavelengths. This then changes the phase vs. voltage response for different wavelengths.

It is also easy to get an idea of the effect of the flyback by repeating the binary phase experiment at different electrode periods. Figure 4.12 shows the results of the binary phase experiment at electrode periods of 2 (where every other electrode is held to a constant voltage and their adjacent neighbors are uniformly changed to some other voltage), 8, 32, and 256 for comparison with Figure 4.10, all at $\lambda = 625$ nm. As expected, for periods of less than 32 electrodes, the binary phase model for the OPA is no longer valid due to flyback. In fact, the results look more like what would be expected from a *sinusoidal* phase grating. The flyback serves to smooth the square wave phase profile to a sine wave profile, as well as damp the phase dynamic range.
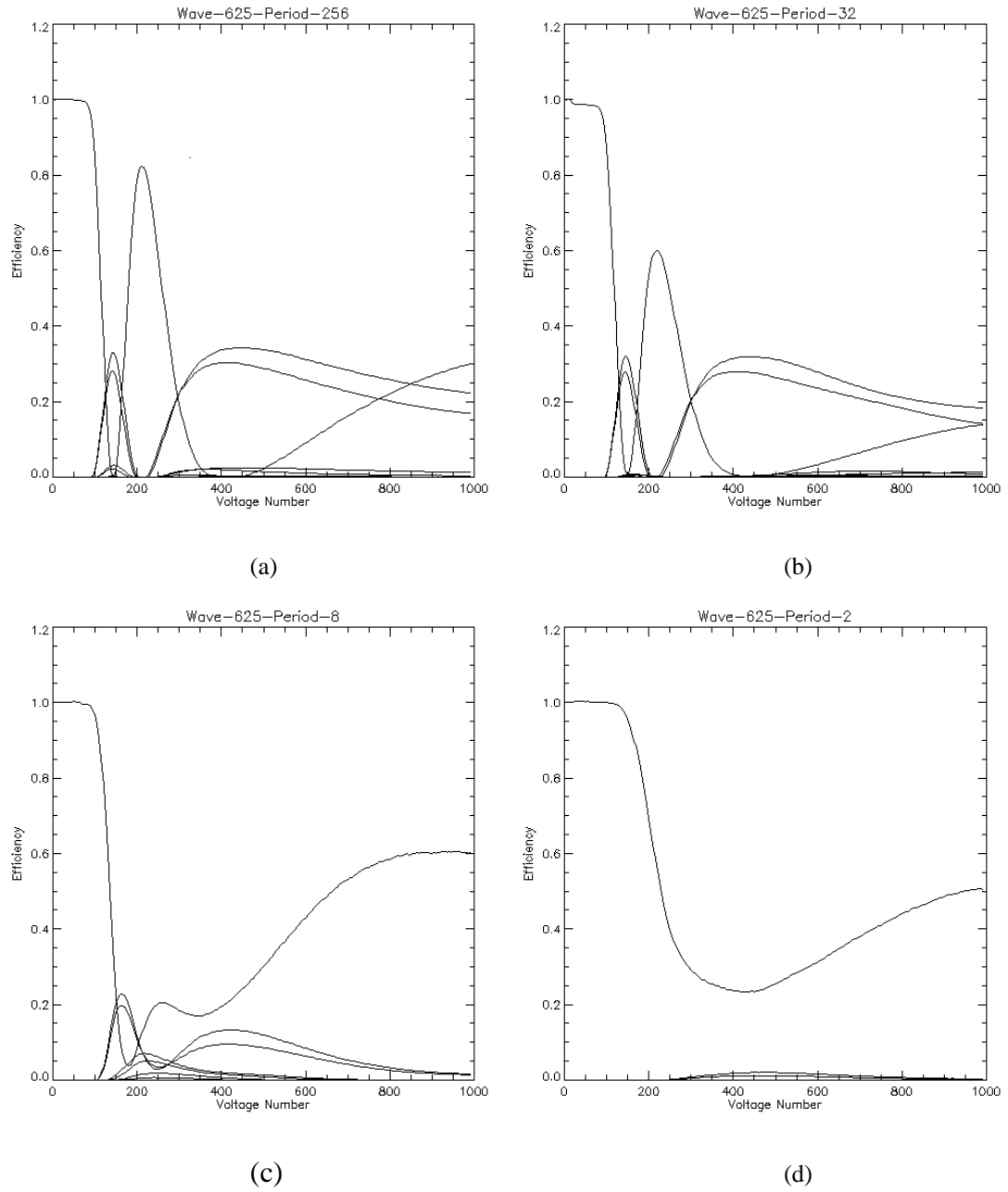
Figure 4.12: The binary phase experiment data at four different OPA electrode periods: (a) 256, (b) 32, (c) 8, and (d) 2.

OPA transfer function

The binary phase experiment has been useful in characterizing the phase response of the OPA and confirming that it behaves as expected. Another interesting aspect of the binary phase experiment is that the resulting diffraction orders *are* the frequency components of the profile that was programmed to the OPA. Perhaps linear systems theory could be utilized to characterize the transfer function, $H(\xi)$, of the OPA when it is programmed as a binary phase grating. The binary phase profile of the OPA can be thought of as a function $f(x)$ with frequency content $F(\xi)$ that is found by taking its Fourier transform. The resulting diffraction pattern in the focal plane of the reimaging lens can be thought of as a function $G(\xi)$ that is the result of the product of $F(\xi)$ with the transfer function of the optical system, i.e., $G(\xi) = H(\xi)F(\xi)$. The diffraction pattern acquired via the experiment is $G(\xi)$, and the transfer function of the OPA is found by calculating $H(\xi) = G(\xi)/F(\xi)$.

An important distinction to be made here is that this is not the optical transfer function (OTF) of the OPA, but rather the transfer function of the 'linear system' as a black box. The situation is depicted in Figure 4.13, where the input/output relationship is a convolution of the input with the point spread function (PSF) of the system. The input is the digital signal sent to the OPA, $f(x)$, and the output is the far-field irradiance, $G(\xi)$.
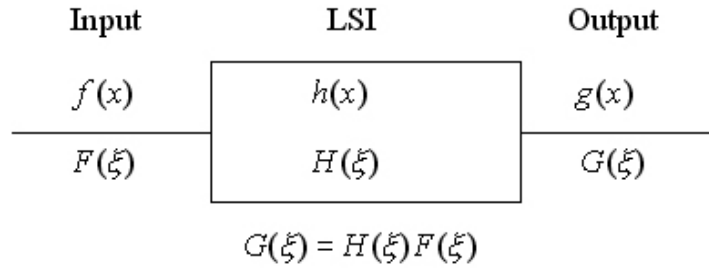
Figure 4.13: Schematic of input/output relationship of a general linear shift invariant (LSI) system.

The reason that the output is G(ξ) rather than g(x) is because of the way the system is set up, which is identical to Figure 5.5(b) in Goodman[22] and is repeated as Figure 4.14 here.
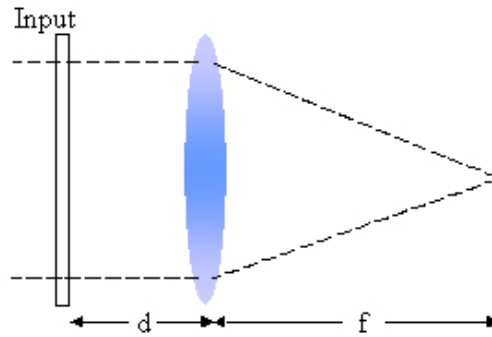
Figure 4.14: Positive lens as a Fourier transformer.[22]

The OPA is modeled as a planar input transparency with transmittance t(x,y). Assuming this input transparency is uniformly illuminated by normally incident, monochromatic plane waves of amplitude A the disturbance just after the OPA using Goodman's notation is:

$$U_l(x, y) = A * t(x, y) \qquad (4.9)$$

The electric field distribution in the back focal plane of the lens, assuming that the Fresnel approximation holds and the spatial extent of the OPA is less than the lens, is:

$$U_f(u,v) = \frac{Ae^{i\frac{k}{2f}\left(1-\frac{d}{f}\right)\left(u^2+v^2\right)}}{i\lambda f} \times \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} t(\xi,\eta)e^{-i\frac{2\pi}{\lambda f}(\xi u+\eta v)} d\xi d\eta \qquad (4.10)$$

where k is $2\pi/\lambda$. The electric field is then the incident light amplitude times a quadratic phase factor times the Fourier transform of the transparency - in this case, the OPA. In the case where the transparency is placed one focal length from the lens, $d = f$, the phase factor disappears and the electric field in the focal plane of the lens is *exactly* a scaled Fourier transform of the OPA transmittance. The diffraction pattern that the CTIS camera acquires, then, is the squared modulus of the frequency spectra of the input, which in the case of the binary phase grating is a binary phase square wave. The digital signal that is sent to the OPA is a square wave – the OPA flyback causes the transparency to be something else by affecting its transfer function, $H(\xi)$. The procedure is to simply divide the square root of the acquired diffraction pattern by the Fourier transform of the input signal to obtain the transfer function of the OPA. It is most convenient to do this using the efficiencies from the binary phase experiment, which are essentially normalized diffraction order weights.

Consider the binary phase experiment at the 625 nm wavelength with a 256 electrode period. If the frequency of the phase square wave is denoted as ξ_o, the diffraction efficiencies at each of the orders are the square of the frequency components at locations n/ξ_o, where n is an integer denoting the order number and ranges from

negative to positive infinity. The theoretical efficiencies are given by equation (4.8) for some binary phase difference m. Now consider the experiment at the same wavelength and binary phase difference m, but with the period halved. The frequency of the square wave is now $\xi_o/2$, and the frequency components are located at $2n/\xi_o$. For all eight possible periods, q = 0 to 7, the frequency components will be located at $2^{(q)}n/\xi_o$, where the square wave period is $256/2^q$. For example, the location of the 3rd order for a square wave with a 2 electrode period (q = 7) falls on the location of the $3*128^{th}$ order of the 256 electrode period (q = 0) square wave. Figure 4.15 shows the locations of the orders in frequency space of the frequency components of the first few input periods. The left column of the figure shows the square wave period. The numbers in the table indicate the order (n) for that period at the frequency location shown by the top row.

	$1/\xi_o$							$8/\xi_o$								$16/\xi_o$
256 q = 1	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
128 q = 2		1		2		3		4		5		6		7		8
64 q = 3				1				2				3				4
32 q = 4								1								2

Figure 4.15: Frequency locations of the diffraction orders for the first few square wave periods.

For the ideal binary phase grating, the even orders do not exist, and the table of Figure 4.15 shows that the odd orders from one particular period overlap with only the even orders of the higher period square waves. Thus the diffraction orders of all eight periods can be superimposed on a single plot. This has been done out to the third order

of the period 2 square wave grating (q = 7). This gives 769 frequency components, from $-384/\xi_o$ to $+384/\xi_o$.

Dividing the diffraction efficiencies of the experimental data at the 625 nm wavelength for all 8 square wave periods by the corresponding efficiencies calculated by equation (4.8) yields the irradiance transfer function (as opposed to the amplitude transfer function) that is shown in Figure 4.16(a). Figure 4.16(a) is derived from the data of Figure 4.16(b) which overlays the experimental efficiencies with the calculated efficiencies. The red lines of Figure 4.16(b) denote the calculated efficiencies with the red asterisks denoting the frequency locations of the first orders of the various square wave periods and the blue asterisks denoting frequency locations of the third orders. The black lines show the detected efficiencies of the corresponding orders at their correct frequency locations. Figures 4.16(a) and (b) only show the frequency components (i.e. diffracted orders), from $-16/\xi_o$ to $+16/\xi_o$. Figure 4.16(c) shows the same transfer function with the range extended to $|64/\xi_o|$ and Figure 4.16(d) shows it again with the full range out to $|384/\xi_o|$. Flyback causes the transfer function to act as a low pass filter so that very little light is coupled to the higher odd orders. This particular transfer function is for electrode number 120, which corresponds to a phase difference in the binary phase experiment of $\varphi = 0.2984\pi$ at the 256 electrode period. This phase difference was chosen for the calculated efficiencies to give a value of 1 for the zero frequency component of the transfer function.

(a)

(b)

(c)

(d)
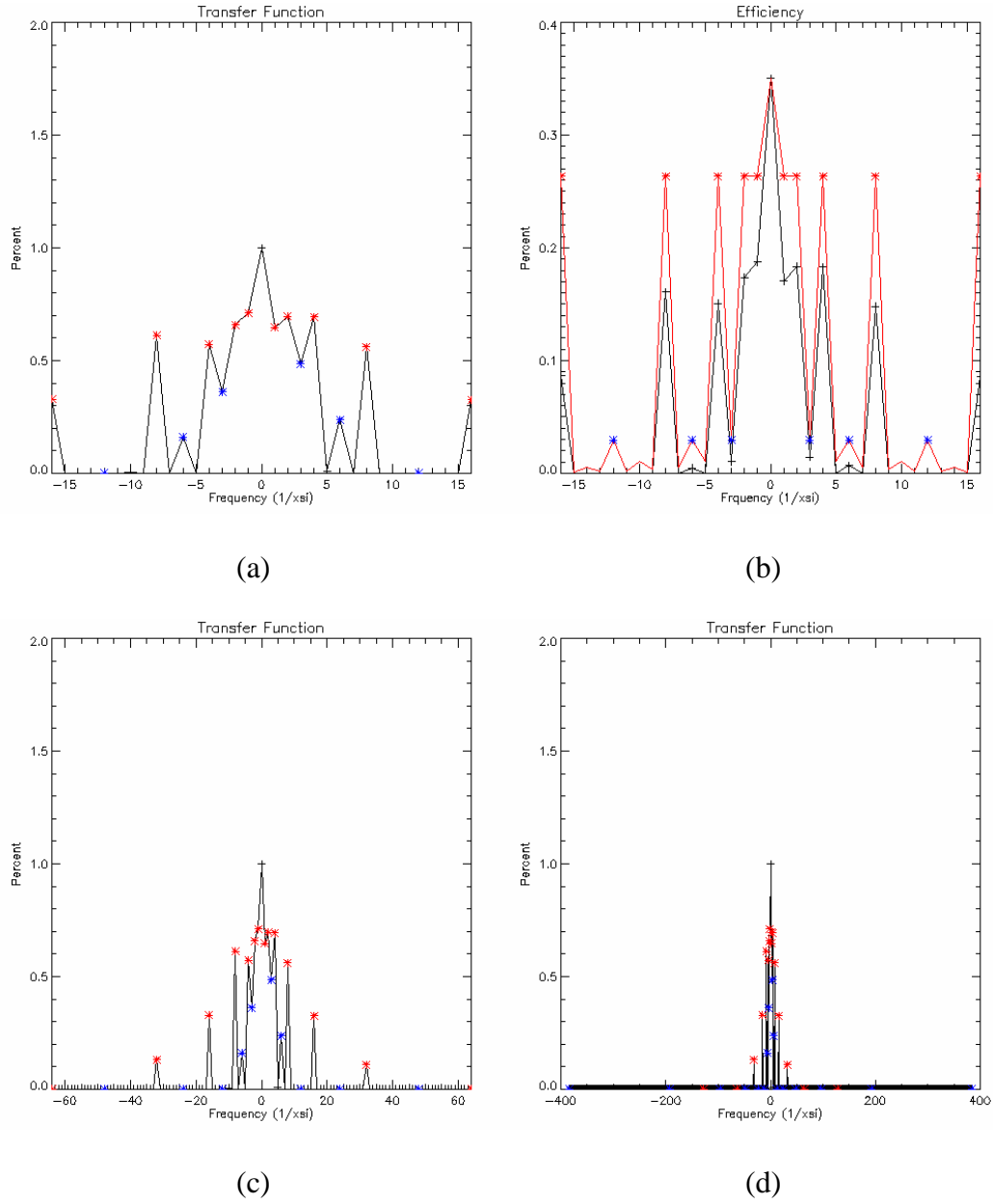
Figure 4.16: (a) The transfer function of the OPA at electrode value 120. (b) The theoretical efficiencies (red) with red asterisks denoting the first orders and blue asterisks denoting the 3rd orders overlaid with the detected efficiencies (black). (c) The transfer function plotted out to 129 components. (d) The transfer function plotted out to 769 frequency components.

(a)

(b)

(c)

(d)

Figure 4.17: (a) Transfer function with electrode value 200 for the lowest 33 frequency components. (b) The same transfer function for all 769 frequency components. (c) The transfer function with 64 frequency components showing where the function exceeds 1. (d) The theoretical and detected efficiencies overlaid showing why certain components of the transfer function exceed 1.

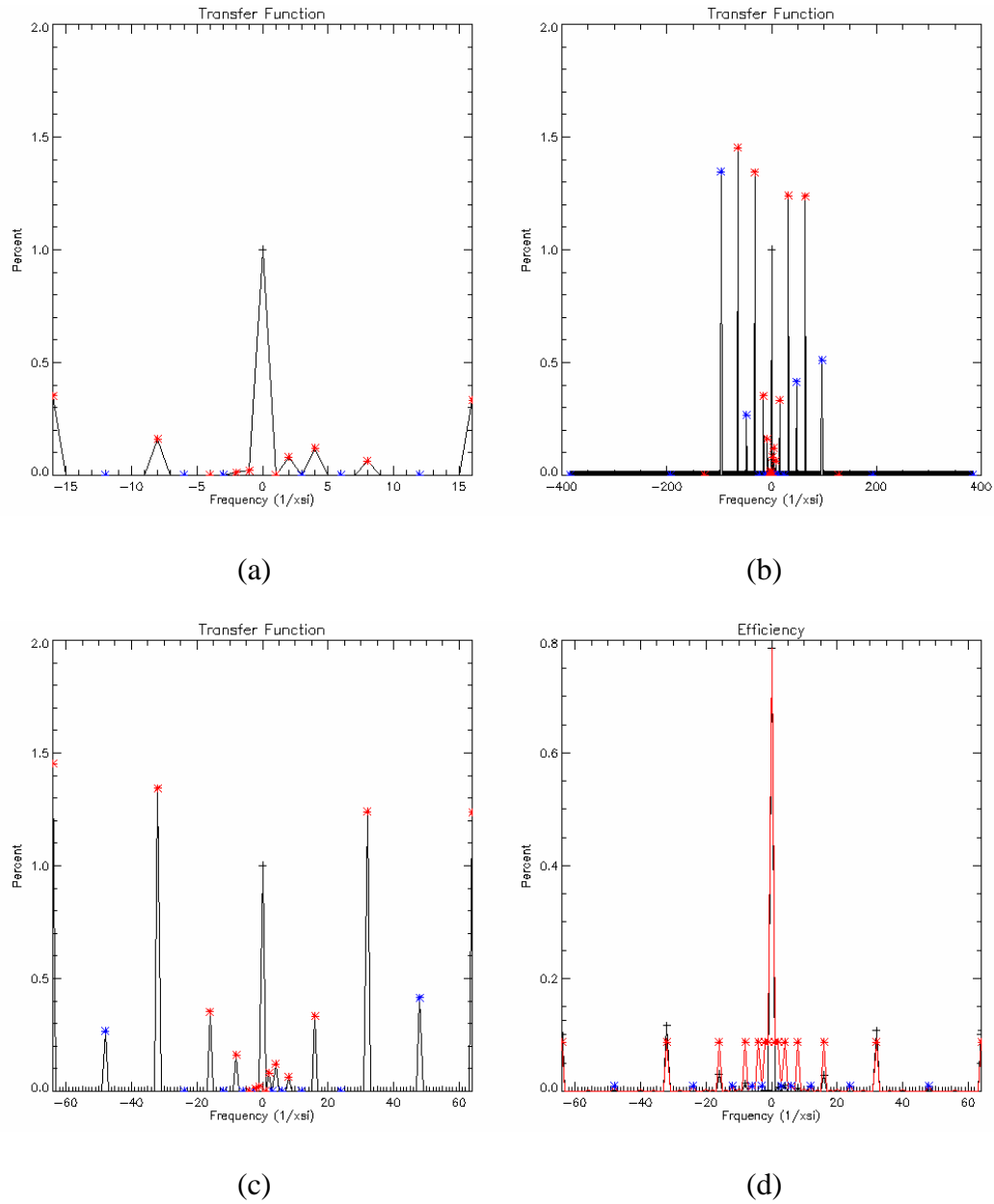The system is, in fact, not linear when all of the available periods are analyzed. To characterize the binary phase experiment of the OPA with a transfer function is a viable approach for electrode periods 256 through 32, but this approach breaks down with the inclusion of the higher frequency square waves. As an example, Figure 4.17 shows the transfer function for a different electrode value, this time at 200, which corresponds to $\varphi = 0.8648\pi$ at the 256 electrode period. Figures 4.17(a) shows the transfer function at the low frequencies, and Figure 4.17(b) shows the transfer function with all of the frequency components that were detected, and some of the components exceed 1! The reason is simply that the theoretical phase value of 0.8648π does not match well for the higher frequency components and the detected efficiencies are divided by a theoretical value that is lower, and thus the transfer function implies gain. Figures 4.17(c) and (d) illustrate the offending components. Figure 4.18 shows this effect in the efficiency vs. voltage plots for electrode periods of 256, 8, 4, and 2. The vertical green lines show the electrode value locations (120 and 200, respectively) for the transfer functions of Figures 16 and 17. Not only does the behavior of the efficiencies change from that of a square wave phase CGH to that of a sin wave phase CGH (a J_o Bessel function rather than a sine-squared function), the amplitude of the CGH wave changes as well. As expected with the flyback, a voltage number of 200 means a smaller peak to valley phase difference at the 2 electrode period than the 256 electrode period because the adjacent values pull the 'high' phase pixels down and the 'low' phase pixels up, thus damping the phase amplitude.
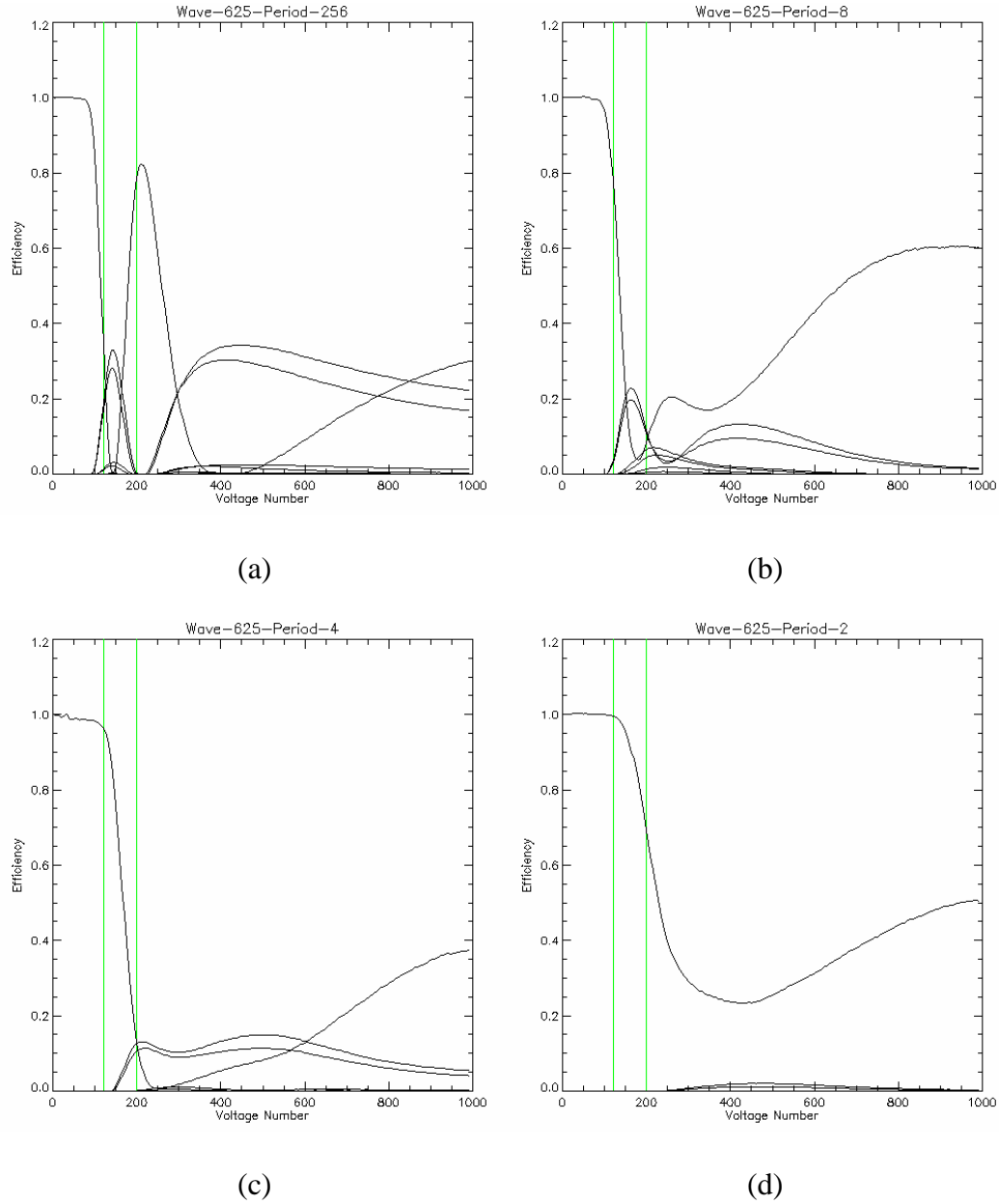
Figure 4.18: Efficiency vs. voltage number plots of the binary phase experiment for periods of 256, 8, 4, and 2. The plot in (a) shows that the efficiency follows a sinusoid squared function that is characteristic of a phase binary square wave and is distorted due to the nonlinear phase response of the OPA. The plot in (d) shows that the efficiency more closely follows a Bessel function that is characteristic of a phase sine wave. Plots (c) and (d) transition between the two states.

In other words, the phase vs. voltage response of the OPA changes for the small period binary phase experiments.

Since the binary phase experiment yields the frequency components of the phase function that actually exists on the OPA, the inverse Fourier transform of the components can be taken to see what kind of function it is. Figures 4.19(a)-(d) show the reconstructed phase profiles from the binary phase experiment for periods 256, 8, 4, and 2 for voltage levels 145, 165, 250, and 400, respectively. The voltage levels were chosen to give a $\pi/2$ phase difference between the two levels for the given period. As expected, the 256 electrode period resembles a square wave, and the 2 electrode period looks like a sine wave with less phase amplitude.

For use as a CTIS disperser, more complicated phase profiles will be needed where the phases between adjacent electrodes must differ. It is not possible to fully characterize the OPA in order to directly program the correct voltages to the OPA to realize a phase profile that is designed a priori in the computer. Fortunately, the CGH disperser design with the OPA can be done experimentally.
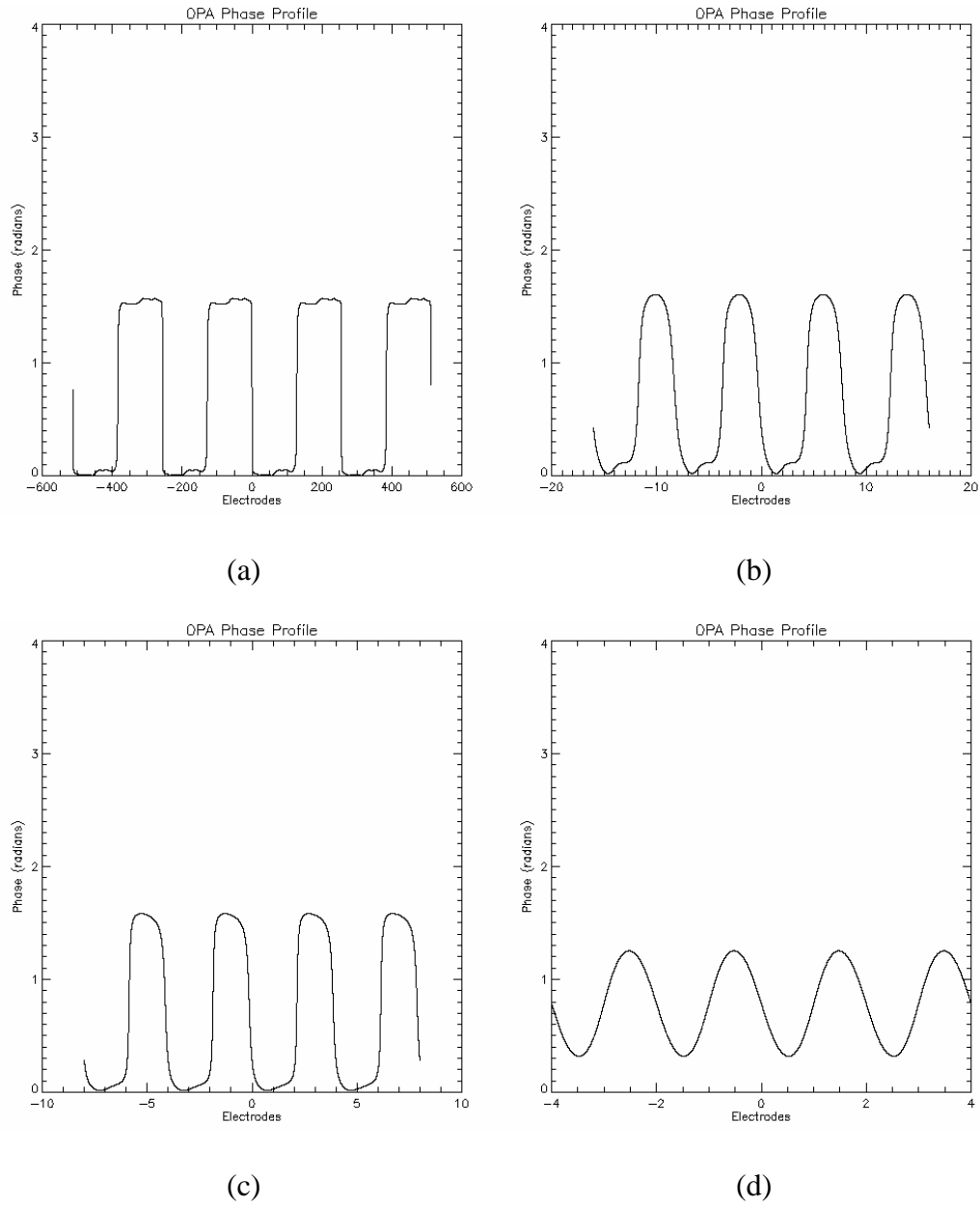
(a)

(b)

(c)

(d)

Figure 4.19: Reconstructed phase profiles from the binary phase experiment for electrode periods of (a) 256, (b) 8, (c) 4, and (d) 2.

Feedback CGH design

The CGH design algorithm used with the OPA is a modified version of the SVD CGH design method that was developed in Chapter 3. When the OPA is used as the CTIS disperser, an 'applied voltage' profile is designed rather than a phase profile. The crux of the SVD CGH design algorithm is that far-field amplitude changes are linearly related to small CGH phase changes. In order to use the algorithm with the OPA, the far-field amplitude changes need to be linearly related to small *applied voltage* changes. The phase as a function of applied voltage is non-linear, but it is a well behaved function (see Figure 4.10(b)) with no discontinuities so it is reasonable that:

$$\frac{dA}{dVolt} = \frac{dA}{d\Phi}\frac{d\Phi}{dVolt} \tag{4.9}$$

The inversion step of the algorithm can be rewritten as:

$$\delta\mathbf{Volt} = -\mathcal{H}^{+}\Delta\mathbf{A}_{e} = -\mathbf{V}\tilde{\mathbf{W}}^{-1}\mathbf{U}^{t}\Delta\mathbf{A}_{e} \tag{4.10}$$

where $\delta\mathbf{Volt}$ is the voltage correction vector (not to be confused with $\mathbf{V}$, which is the matrix of *voltage* perturbation modes). The sensitivity matrix is constructed using amplitude change to voltage change ratios rather than amplitude change to phase change ratios.

This adaptation to the SVD CGH design is referred to as the 'feedback' CGH design algorithm and the experimental setup is once again very similar to that of CTIS calibration. A spatial/spectral point source is placed on axis in the plane of the field stop using the calibration equipment. The main difference between the feedback design and calibration is that with the feedback design, the reimaging lens focal length needs to be

adjusted. Typically this focal length is set to maximize the footprint of the desired orders

on the array and all of the higher orders fall off of the array. For the algorithm, these

higher orders are needed and it is simplest to just shorten the focal length of the

reimaging lens accordingly. For example, for a 1D design that uses 8 electrodes in the

unit cell, eight orders (usually from –4 to +3) will need to be collected.

The three major modifications to the SVD CGH design algorithm for use in

feedback with the OPA are (1) thresholding the smallest few singular values, (2) damping

the voltage correction vector, and (3) strictly constraining the voltage range of the voltage

correction vector. As discussed at the end of Chapter 3, thresholding of the smallest few

singular values is necessary due to nonlinearities, especially with access only to far-field

amplitudes and not the complex electric fields. This thresholding eliminates the wildly

varying elements of the voltage correction vector. Damping the voltage correction vector

imposes a global constraint that causes the algorithm to approach a solution slowly and in

a stable manner. For the damping, the voltage correction vector was divided by 7. This

number is chosen based on experience to limit the range of voltage correction to less then

200 (20% of the dynamic range). These two modifications affect the stability of the

algorithm's convergence, though it is entirely possible that the voltage correction vector

will still cause some of the applied voltages to be corrected outside the possible range.

In the conventional SVD CGH design algorithm in the computer, the dynamic

range of phases is inconsequential because at any particular phasel a multiples of 2π at

the design wavelength can be subtracted with little effect to the amplitudes in the

irradiance plane. If the corrected voltage numbers to be applied the OPA extends above

1023 or below 0, it is unknown what voltage number to subtract (or add) that corresponds to a multiple of 2π phase. Therefore, the offending elements of the voltage correction vector are reset to the center of the voltage dynamic range and it is left to the next iteration to adjust them to the correct value. These modifications have been successful and the feedback design converges to a reasonable CGH.

The two crossed LC cells of the OPA are denoted as channel 5 for the horizontal electrodes (vertical *orders*) and channel 6 for the vertical electrodes (horizontal *orders*), which is how they are denoted in Raytheon's software. The results of the feedback CGH design with the same goal as the 1D example of Chapter 3 was performed and the results for channel 6, the vertical electrodes, are shown in Figure 4.20 below. Figure 4.20(a) shows the *relative* efficiencies of the diffracted orders. The relative efficiencies are calculated by dividing each order by the total light that falls on the array. Figure 4.20(b) shows the absolute efficiencies, which are calculated by using the *incident* light that is acquired when the OPA is removed. Figure 4.21 shows the reason for the radical difference between the relative and absolute efficiencies of the OPA, which is the spectral transmission of the OPA. Because of the AR coating, the OPA has a steep spectral transmission curve that dominates the absolute efficiencies. The relative efficiencies give a better assessment of the performance of the feedback design, which compares well with the etched CGH of Figure 3.12. If both channel 5 and channel 6 of the OPA are programmed to have the response shown in Figure 4.20, the instrument will be a 5 x 5 CTIS, meaning it will have a grid of 5 x 5 projections from which to reconstruct the data cube and is typical of many of the CTIS instruments that have been made in the lab.
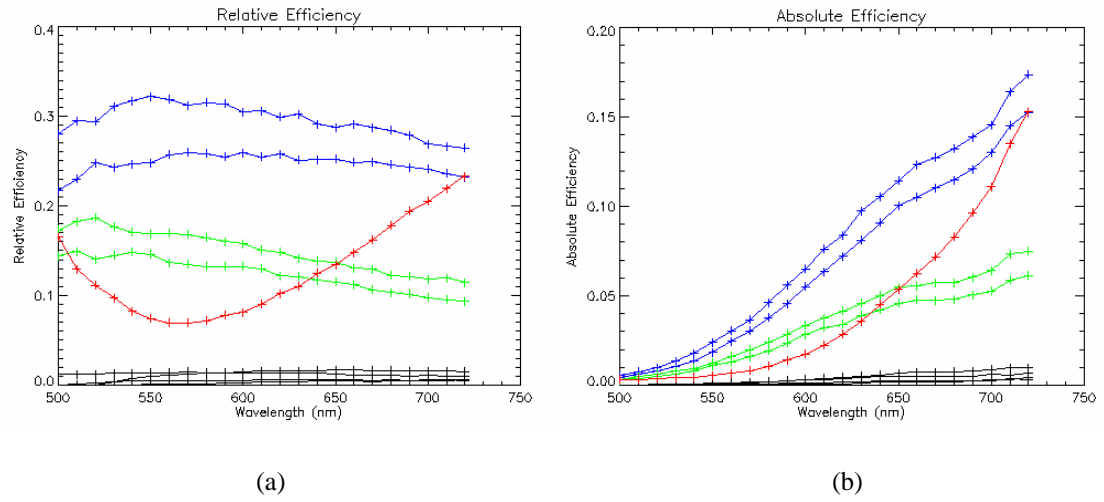
Figure 4.20: (a) Relative efficiencies of the OPA and (b) absolute efficiencies of the OPA. The zero order is plotted in red, the first orders green, and the third orders blue.
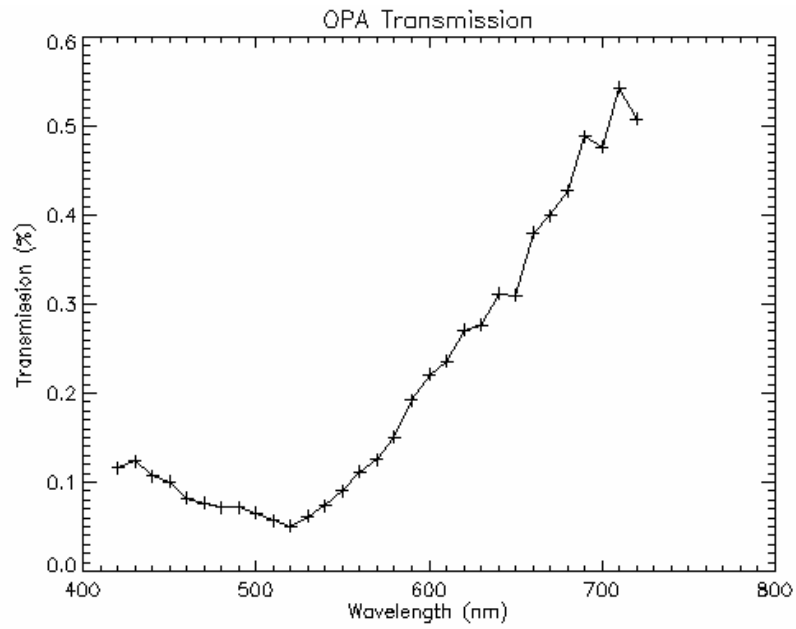

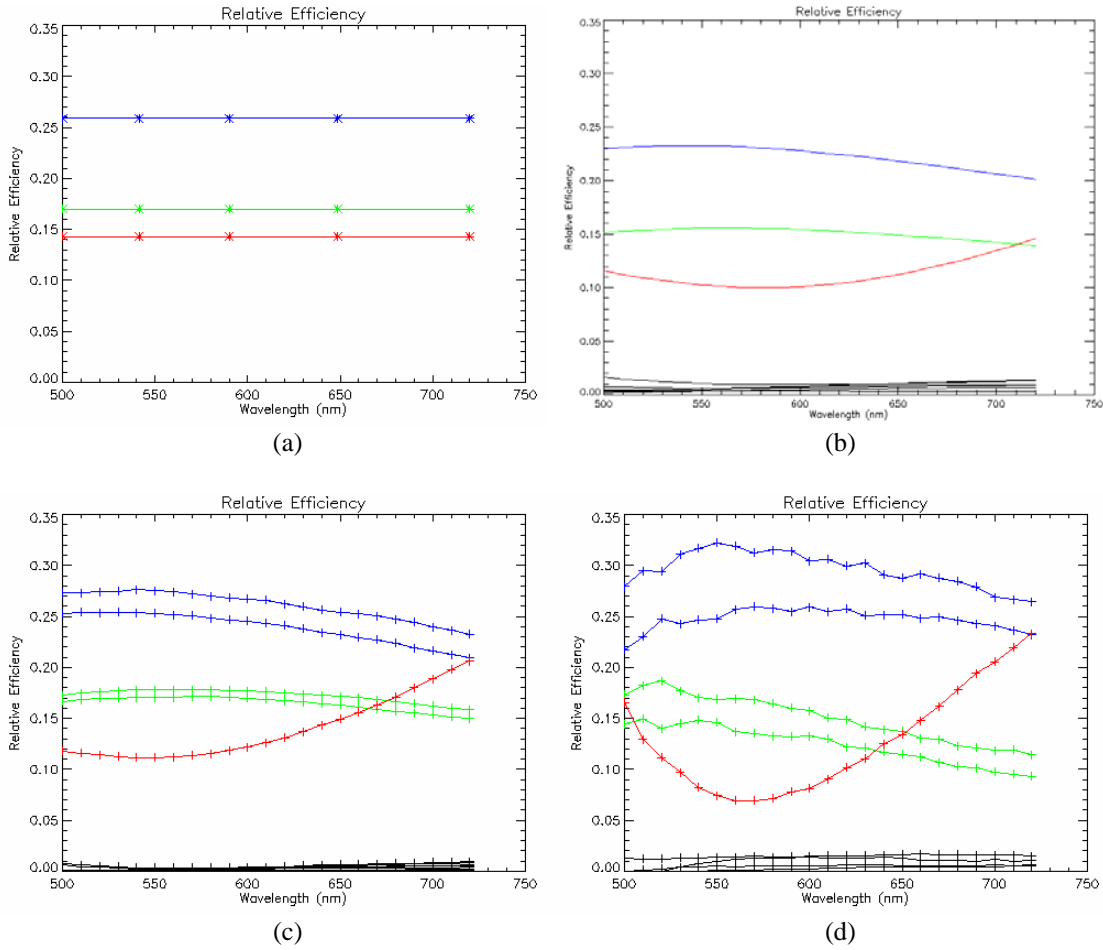**Figure 4.21:** Transmission of the OPA.

Figure 4.22: Comparison of the efficiency plots of (a) the CGH design goal, (b) simulation, (c) measured relative efficiencies of the etched CGH, and (d) measured relative efficiencies of the OPA CGH design.

Figures 4.22(a)-(d) show the spectral efficiency plots of the CGH design goal, simulation, measured relative efficiencies of the etched CGH, and measured relative efficiencies of the OPA CGH design for comparison. Note the powerful advantage of the feedback CGH design in 'real time' with the OPA. Figure 4.22(d) can be achieved in a day, whereas Figure 4.22(c) constitutes shipping a CGH design to a willing and competent vendor, paying for it, waiting for it to be shipped back, and testing it to see whether or not

it worked. After testing, provided the CGH performs well (JPL has continued to etch excellent CGHs), you can't change it. Besides the practical factors listed above, another major advantage of the feedback CGH design is that the design is changed according the *measured* results.

The convergence of the feedback algorithm can also be plotted for comparison

(a) $\left| \delta \overline{V} \right|$

(b) $\sigma_{\delta V}$

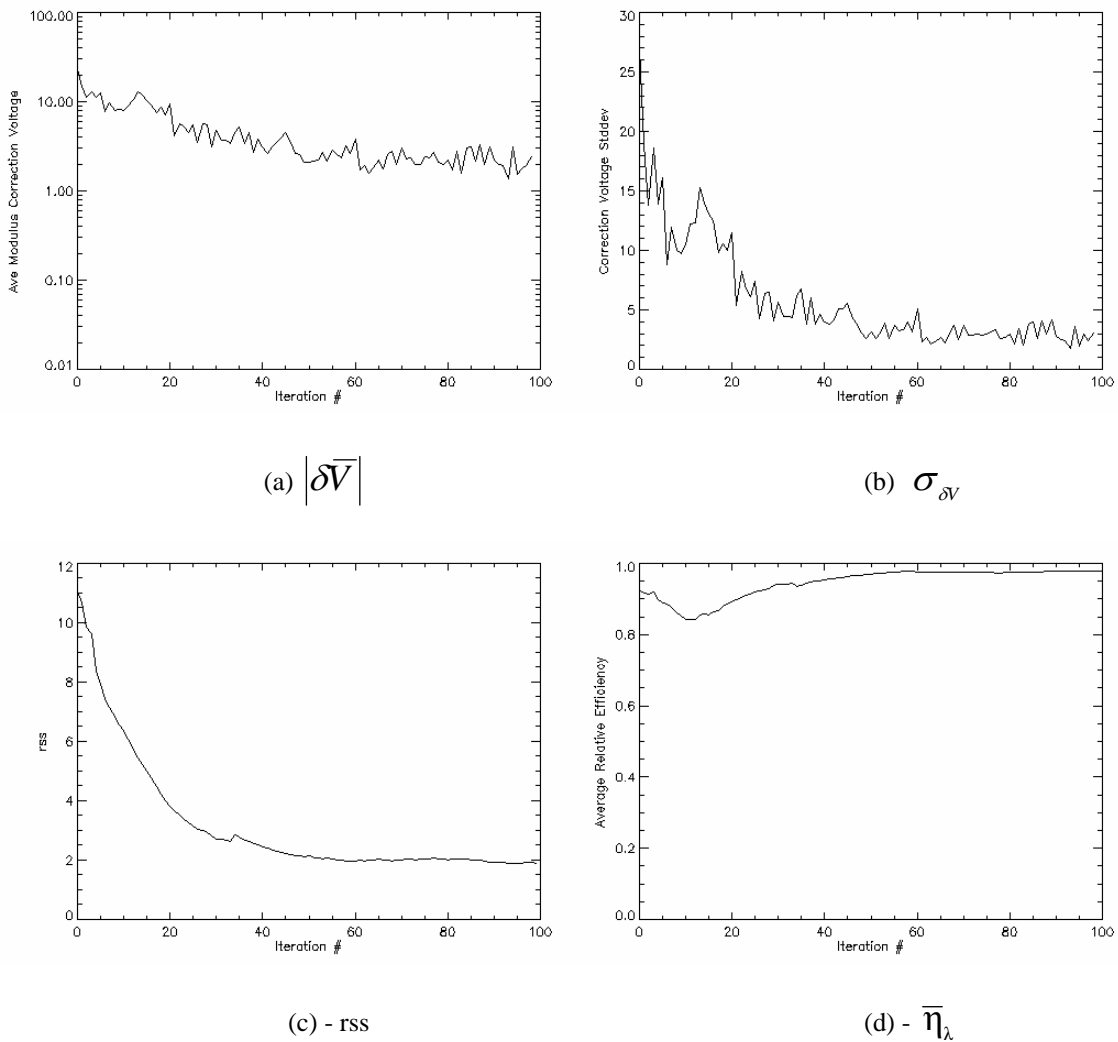(c) - rss

(d) - $\overline{\eta}_{\lambda}$

Figure 4.23: (a) Average modulus of the voltage correction vector, (b) standard deviation of the voltage correction vector, (c) rss of the design, and (d) $\overline{\eta}_{\lambda}$ of the design.

with Figure 3.12 in the convergence section of Chapter 3. The plot of the average

modulus of the voltage correction vector, the average variance of the voltage correction

vector, the rss, and the spectral average of the efficiency in all of the *desired* orders as a

function of iteration number are shown in Figure 4.23.

The comparison between the OPA design of Figures 4.23(a) and (b) and the CGH

design convergence of Figures 3.12(a) and (b) are not direct – the former depends on a

voltage correction vector in units of ADUs, and the latter depends on a phase correction

vector in units of radians. They show that the stability of the OPA algorithm is analogous

to version 3 in Figure 3.12, where there is thresholding of only the smallest singular value

in the simulation. Two things to take note of are the differences between the simulation

results in Figures 3.12(c) and (d) and the actual design with the OPA in Figures 4.23(c)

and (d). First, the root-sum-squares, or rss, which measures the match to the desired goal

pattern takes more iterations to settle to a minimum value in the feedback design with the

OPA. This is a manifestation of the damping of the voltage correction vector. Second,

the efficiency in the desired pattern starts off very high in the feedback design with the

OPA. This is because nearly all the light is in the zero order at that iteration, and the

efficiency drops off as light is diffracted to higher orders as the voltage vector is

corrected. The efficiency rises once again as the light that is taken from the zero order is

actually put into the *correct* higher orders that correspond with the desired diffraction

pattern.

Voltage dependent scatter

The results of the feedback design in the previous section are presented for only

channel 6 of the Raytheon OPA, which is the vertical electrode LC cell. The reason for

this is that the other cell, channel 5, exhibits behavior that appears to be 'voltage

dependent scatter.' Through most of the dynamic phase range of this LC cell

(corresponding to voltage numbers between 0 and 300) a point source placed on axis in

the field stop will be scattered by the OPA to image to a smear of light on the FPA.

Figure 4.24 below shows images (in reverse contrast) of this scattering for voltage values

of 0, 50, 100, 150, 200, and 300. All electrodes are changed to the same voltage number,

and channel 6 is held to a constant voltage number of 0. The scattering is discrete in

nature with a very fine spacing between the points (thus perhaps 'voltage dependent

grating ghosts' would be a more appropriate term). This leads to the conclusion that

there is some periodic error in the voltages that are applied to the electrodes. It seems as

though for small voltage values, each group of 256 electrodes across the OPA have

different voltages leading to a large period grating effect. Raytheon is of the opinion that

there may be corrosion of the electrodes that cause the voltages to differ between groups

of electrodes for small voltage values.[23]

This voltage dependent scatter has thus far thwarted attempts to perform the

feedback CGH design with this LC cell. Consequently, the implementation of tunability

in the next chapter is performed in 1D varying only the vertical electrode LC cell while

holding this horizontal electrode cell to a high constant voltage level at which the

scattering is a minimum. Diffraction in the vertical direction is achieved by using one of the 1D etched CGHs that are described in Chapter 3.
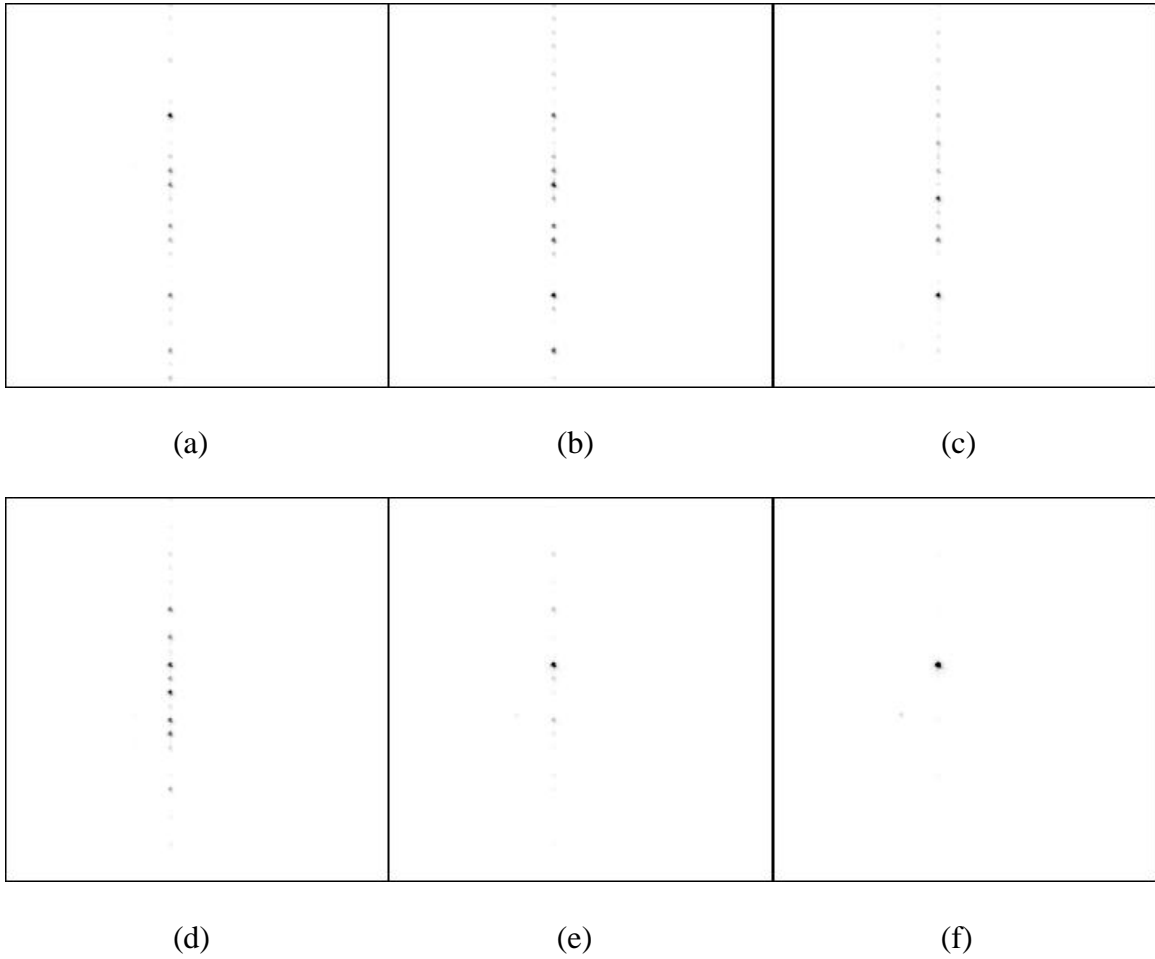

Figure 4.24: CTIS images of a point source as channel 5 (the horizontal electrode LC cell) of the OPA is changed to voltage numbers (a) 0, (b) 50, (c) 100, (d) 150, (e) 200, and (f) 300.

CHAPTER 5 – TUNABLE CTIS

Tunable CTIS parameters

Given a successful demonstration of the Raytheon OPA as a tunable CTIS disperser from Chapter 4, the subject of this chapter is the implementation of that tunability. The electronically tunable disperser allows three of the CTIS parameters to be altered: (1) the grating period of the disperser, (2) the number of diffraction orders (i.e. reconstruction projections), and (3) the efficiency of the orders as a function of wavelength.

Changing the first parameter, the grating period, changes the dispersion across the CTIS FPA, and thus the spectral resolution limit of the instrument can be altered. However, this may cause some orders to fall off the array if the grating period is too small. Conversely, if the grating period is too large the FPA is being used inefficiently and there is also the possibility that the first orders will start to overlap with the zero order, which is unacceptable for reconstruction.

Changing the second parameter, the number of diffraction orders on the FPA, amounts to trading off reconstruction quality for reconstruction time and signal to noise ratio, both of which depend on the number of projections from which the data cube is reconstructed. An example of combining these first two parameters is the case where more signal to noise ratio is desired for a CTIS with a 5 x 5 grid of diffraction orders without reducing the spectral resolution of the reconstructions. The light in the second orders is coupled to the first orders so that there is now a 3 x 3 grid of diffraction orders,

and the grating period is halved so that the dispersion of the first orders equals that of the second orders in the original 5 x 5 configuration.

The most interesting, and difficult, parameter to change is the third parameter, the efficiencies of the diffracted orders as a function of wavelength. An example of the utility of this parameter is to design the disperser such that the efficiency of a particular diffraction order as a function of wavelength is the inverse of the background of a scene, and thus the background is 'whitened' in that order. If there is an object in the field of view that has a spectra other than that of the background, it can be identified directly from the raw CTIS image before reconstruction. Other things have been suggested as well, such as programming a spectral modulation to the orders that may spread information into the frequency region of the missing cone (see Chapter 3), or perhaps something analogous to phase shifting interferometry[24] could be done by taking several images with different disperser configurations. In this case, the thought is that the images could be used to determine the 3D data cube without using the current MART and EM reconstruction algorithms.

Experimentally, it is difficult to demonstrate tunability by changing the spectral efficiencies of the diffraction orders using a liquid crystal spatial light modulator. In fact, it may be impossible with the borrowed OPA that was not designed this project. The lack of sufficient phase range (up to several waves), the low transmission in the visible wavelength region, the voltage dependent scatter mentioned at the end of Chapter 4, and the x-y separability imposed by the 1D architecture that reduces the degrees of freedom

in the design are the main practical factors that prevent experimentation of this parameter using the OPA.

It was chosen to demonstrate CTIS tunability by changing the second parameter, the number of diffraction orders. The goal is to show that the spectral resolution limit of the instrument, the signal to noise ratio of the orders, and the reconstruction time are all 'tunable' and to denote the tradeoffs by simply changing the number of diffraction orders. It is also an interesting experiment because the resolution limit of the instrument is, at present, defined only by a geometric calculation. A spreadsheet is used to calculate the dispersion based on perfect geometric imaging, shown for the tunable CTIS in Figure 5.1.

Figure 5.1: Spreadsheet calculating the geometric properties of the CTIS instrument.

The red arrow in Figure 5.1 points to the calculated dispersion across a single FPA pixel

with a 5 x 5 diffraction pattern, which happens to be roughly 1 nm in the highest orders

(in this case the ring of 16 second orders are the highest orders). The dispersion is

calculated by the grating equation:

$$\sin(\theta) = \frac{m\lambda}{\Lambda} \approx \frac{d}{f}$$

$$\sin(\theta) \approx \tan(\theta) = \frac{d}{f} \tag{5.1}$$

$$\Delta\lambda = \frac{\Delta d\Lambda}{mf}$$

where Δd is the width of a FPA pixel in the direction of the dispersion, Λ is the grating

period, f is the focal length of the reimaging lens, m is the order number, $\Delta\lambda$ is the

dispersion, and the small angle approximation is used. Keeping in mind the Nyquist

critical sampling rate which says that there must be at least two samples (i.e. two pixels)

in order to resolve a feature, a reasonable inference is that the spectral resolution limit of

the tunable CTIS should be 2 nm, which is twice the dispersion across a single pixel in

the highest orders. The spectral sampling of the CTIS is determined by its calibration,

and the CTIS is rarely calibrated to its geometric limit that is determined by the

dispersion calculation. The first part of the experiment in this chapter will be to develop

a calibration with sufficient spectral sampling.

The reason that the term 'spectral resolution limit' is used is because the spectral

resolution of the CTIS is coupled to the spatial frequency content of the raw image and is

optimum for images containing high spatial frequency content, i.e. images of point

sources. The reason for this, as mentioned in Chapter 2, is because the CTIS is based on

limited angle tomography and high spectral frequency content is lost in the missing cone (see Figure 2.2) for sources having only low spatial frequencies. Thus the metric that is chosen here is the best possible spectral resolution, which occurs with images of point sources, and this is denoted as the *spectral resolution limit* of the CTIS.

It is hypothesized that in the event that the optics of the instrument are corrected well enough such that a point source images to a single FPA pixel, this geometric spectral resolution limit that is calculated by the grating equation could be reached experimentally. The following section presents an experimental method for determining the spectral resolution limit of the CTIS after reconstruction.

Spectral resolution experiment

The experiment begins with a definition of resolution that can be found in Born and Wolf[25] and that was first proposed by Lord Rayleigh.[26] Suppose that light entering a Fabry-Perot spectrometer is composed of only two quasi-monochromatic components of equal irradiance. The components are said to be 'just resolved' if their respective maxima overlap with the corresponding minima of the other component, as depicted in Figure 5.2. The experimental definition of the CTIS spectral resolution that is chosen here uses Lord Rayleigh's criterion *after* reconstruction.
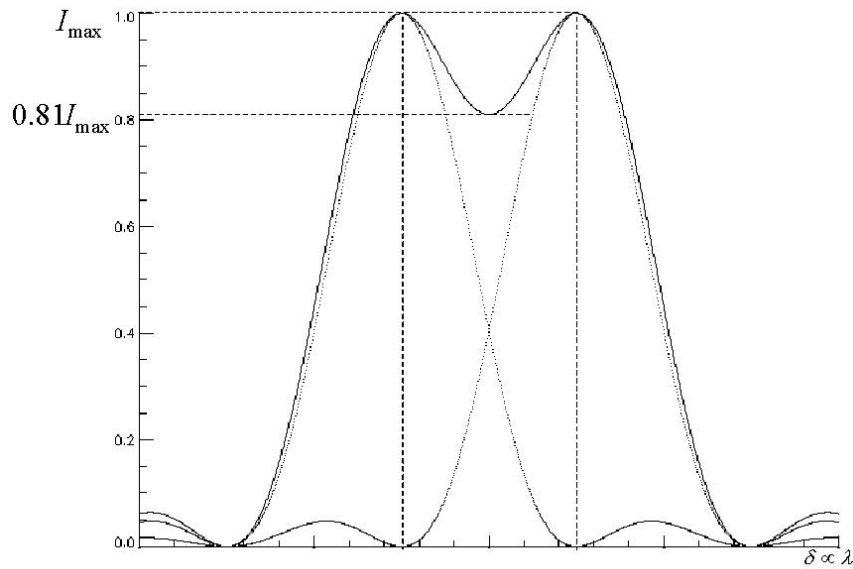
Figure 5.2: Illustration of Lord Rayleigh's criterion for a Fabry-Perot spectrometer for two 'just resolved' quasi-monochromatic components.[25]

The CTIS resolution is defined by the resolvability of two monochromatic point sources after reconstruction. Notice that if truly monochromatic spectral components could be passed through the Fabry-Perot without degradation (perfect imaging) and could be acquired continuously (continuous sampling), the spectral resolution would be infinite because it is the components' spectral widths that add to masquerade as a single component at the average wavelength. There are three main factors that degrade the spectral resolution of a spectrometer: (1) the finite spectral bandwidth of the source used to calibrate the instrument, (2) the optics, and (3) the sampling of the final image. A different metric, the full width at half max of a single spectral component (FWHM), could also be used. The FWHM metric depends on the symmetry of the spectral

components and becomes ambiguous when they are asymmetric. The sources used for this experiment use symmetric spectral components. However, when the sampling is pushed to the limit of the FPA there are errors that crop up in the components' locations that translate to asymmetric reconstructions. Appendix B discusses the FWHM metric and presents a few more reconstruction results than are presented in the text of this chapter.

The spectral resolution of CTIS instruments depends on the spatial distribution and spectral bandpass of the sources used for calibration. Typically it is desired to use a source which images to a single pixel. This is rarely the case and there is usually a spread in the calibration 'point' source images as shown in Figure 2.6. For the present experiment, the calibration of the tunable CTIS uses sources with a relatively large spatial spread (due to the use of a 400 μm diameter fiber) and a large spectral bandwidth, $\Delta\lambda$, from the monochromator, both of which are large in order to get appreciable signal. The current situation, then, is that the limit to the spectral resolution imposed by the instrument itself cannot be reached due to insufficient calibration sources. The solution to this problem, presented in the next section, will be to *simulate* sufficient calibration *images* using the *sources* that are available. The remainder of this section describes the conventional calibration of the tunable CTIS before the simulation.

The calibration of the tunable CTIS for the spectral resolution experiment was performed from 640 nm to 660 nm with a sample every 0.2 nm, each of which had a 5 nm FWHM spectral bandwidth, all located on-axis. Therefore, there are 101 calibration images, or spectral bands, for each pixel in the field stop. The source was then moved

off-axis and 51 images were taken from 645 nm to 655 nm in steps of 0.2 nm. This off-axis data set serves as the object data for the experiment. To obtain an image of an object with two spectral components, any two images from the object data set were simply added together. The convention that was chosen was to always use the 645 nm image (i.e. spectral component) and add another image at a higher wavelength. A plot of the 645-655 nm dual component object source spectra after reconstruction is shown in Figure 5.3(b). Figure 5.3(a) shows the raw CTIS image in reverse contrast and processed to improve the visibility of the orders. Figure 5.3(c) shows a zoomed in view of the zero order image of the 'point' object, which is clearly not a point at all and will be referred to as the 'object source image' of the 'object source'. The plot in Figure 5.3(b) is along the spectral axis of the reconstructed data cube at the object source image spatial location, and Figure 5.4 shows the spectral images (or slices) of the reconstructed data cube from 640 nm to 659 nm in 1 nm steps. Figure 5.3(b) clearly shows that the two spectral components of the object source are resolved.
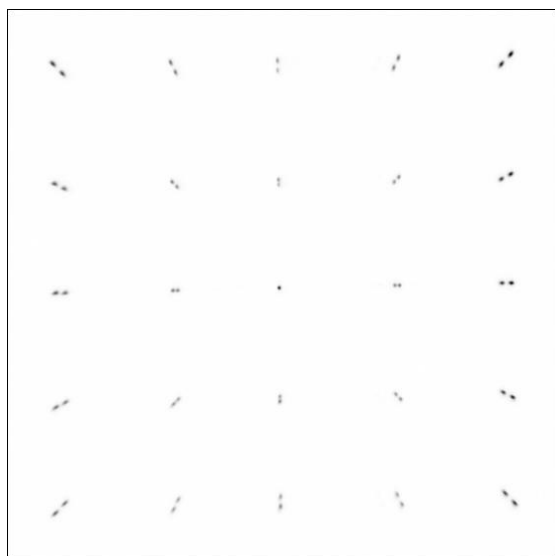
(a)
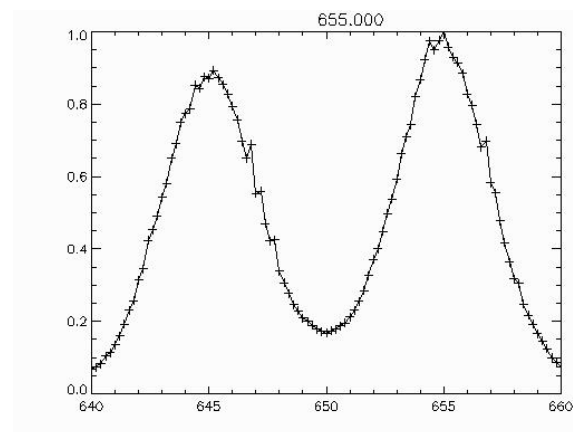
(b)
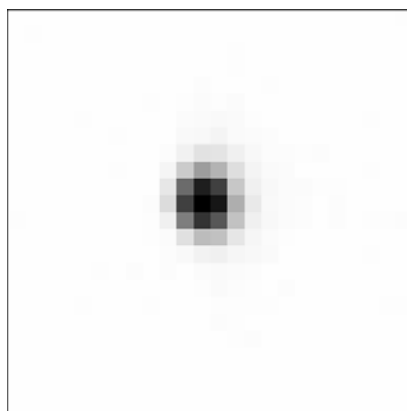
(c)

Figure 5.3: (a) CTIS raw image of the 645-655 nm dual component object source, (b) the spectra of the reconstructed object source, and (c) a zoomed in view of the central 'point' object.
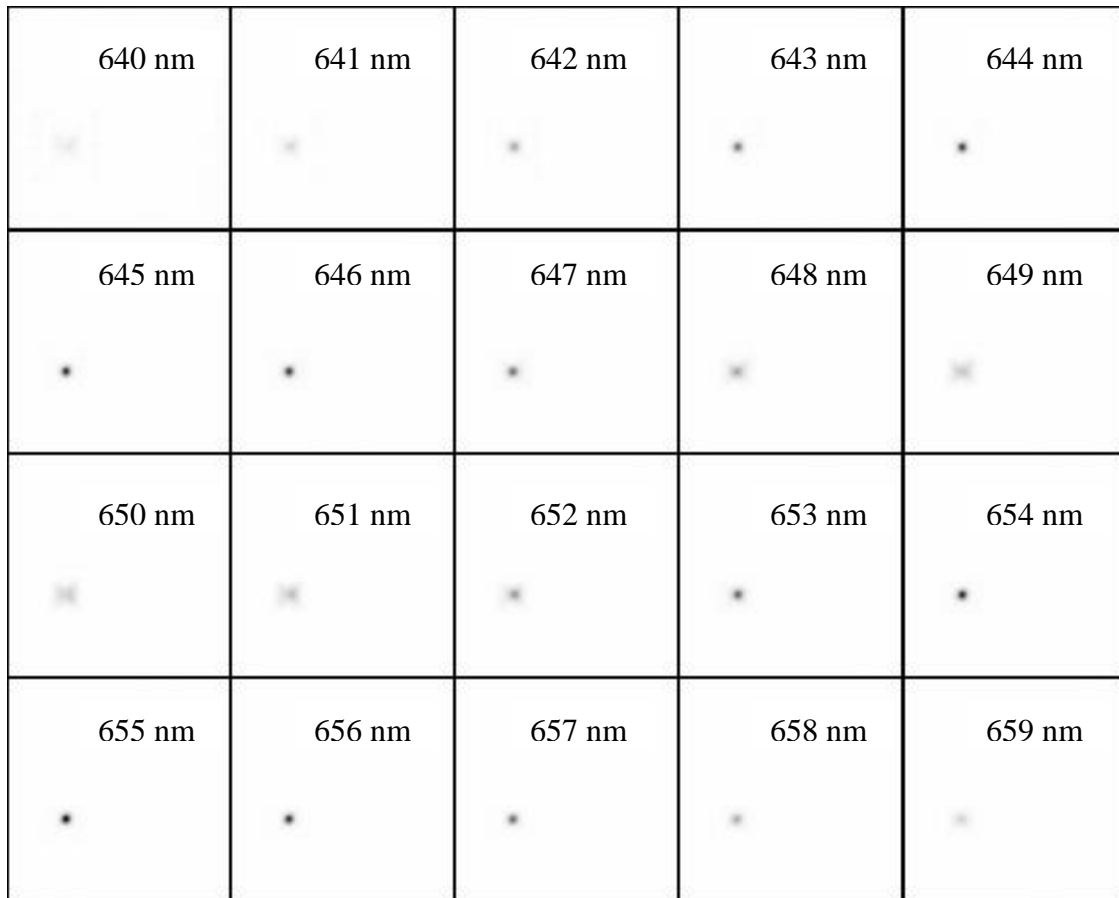
640 nm	641 nm	642 nm	643 nm	644 nm
645 nm	646 nm	647 nm	648 nm	649 nm
650 nm	651 nm	652 nm	653 nm	654 nm
655 nm	656 nm	657 nm	658 nm	659 nm

Figure 5.4: Shown from right to left, top to bottom are the spectral images of the reconstructions at 1 nm intervals from 640 nm to 660 nm.

Reconstructions are repeated for a 645-649 nm object source image and a 645-651 nm object source image. Figures 5.5(a)-(b) show the spectral plots of the reconstructions from these object source images along side the spectral plot of the 645-655 object source image in Figure 5.5(c). Figure 5.5(a) shows the reconstructed spectra of the 645-649 nm object source and demonstrates that the current sources in the lab are not sufficient for calibration or use as object sources to experimentally determine the spectral resolution

limit of the tunable CTIS. Figure 5.5(b) shows that with the current sources, it would be

inferred that the CTIS is just able to resolve two components separated by 6 nm, and

therefore the spectral resolution is 6 nm, which is three times the geometrically calculated

limit. This is not the spectral resolution limit, and a method for providing a quasi-
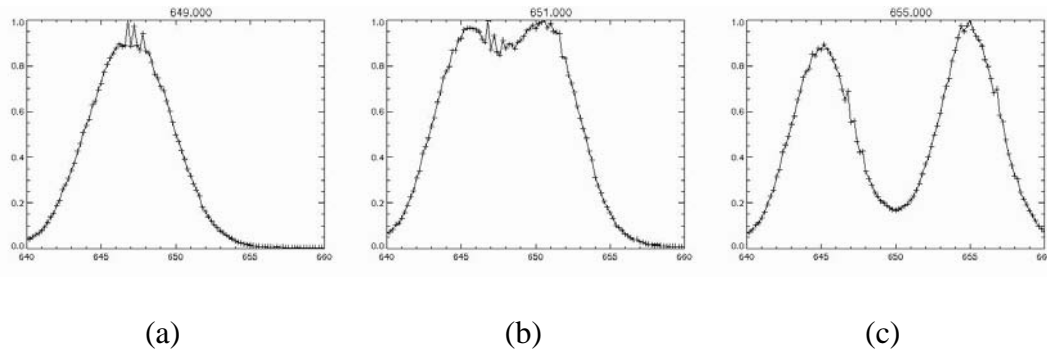
monochromatic point source is needed to continue.

| (a) | (b) | (c) |

Figure 5.5: The spectra at the reconstructed point for the (a) 645-649 nm, (e) 645-651 nm, and (f) 645-655

nm components respectively.

Point source simulation

      Due to the low transmission of the OPA and the relative small throughput of the

tunable CTIS, a large core diameter fiber was needed to put enough signal in the plane of

the field stop to be detectable by the FPA. A point-like source will therefore not be

available for this experiment. However, the data that was taken can be used to *simulate* a

point source with a quasi-monochromatic bandpass. This is done by centroiding each of

the diffracted orders in the calibration and object images. Only three reasonable

assumptions need to be made for centroiding to be a valid simulation of a quasi-

monochromatic point source for the CTIS: (1) the source must have a circularly symmetric irradiance distribution, (2) the CTIS PSF must be circularly symmetric, and (3) the spectral content of the source must be symmetric about its center wavelength. If these assumptions are true, centroiding the original data is a good approximation of a quasi-monochromatic point source.
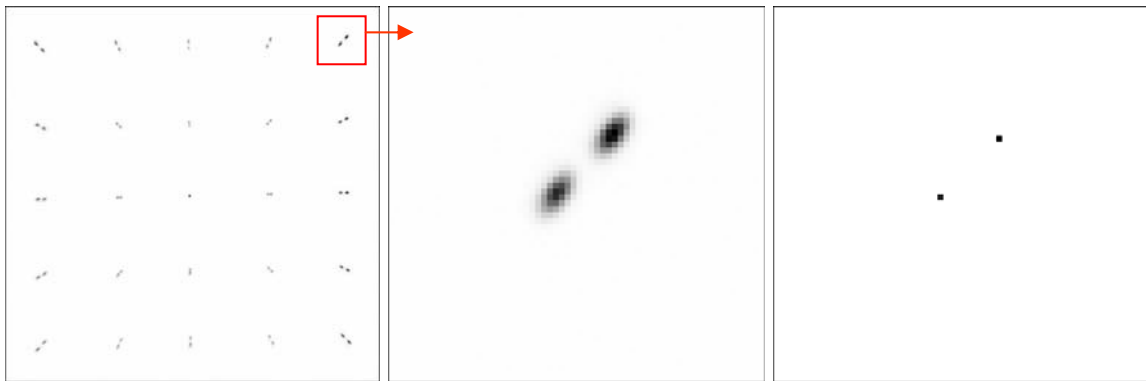

Figure 5.6: (a) The raw CTIS image, (b) an expanded view of the upper right diffraction order, and (c) the centroid of the upper right diffraction order.

Figure 5.6(c) shows the centroids of the upper right diffraction order of the raw CTIS image of Figure 5.3(a).

The reconstructions are repeated after creating simulated calibration and dual component object point source images with quasi-monochromatic bandpass using centroiding. The spectral plot of the results of the reconstruction of the 645-646.6 nm dual component object point source image is shown in Figure 5.7. This is the result before convolving the diffracted orders with the CTIS PSF which must be accounted for

and will be considered in the next section. The spectral resolution is now 1.6 nm, which

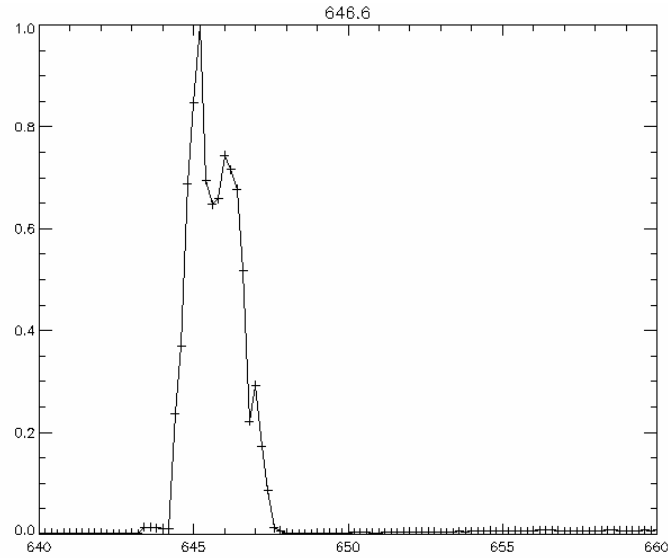exceeds the resolution predicted by the geometric dispersion calculation.

Figure 5.7: Reconstructed spectra of the simulated 645-646.6 nm dual component point source image.

The centroiding results in continuous spatial locations for the diffracted orders, so

it is tempting to think that the sampling is now 0.2 nm and it should be possible to resolve

a dual component object point source image with its components separated by 0.4 nm. In

order to reconstruct, however, the sampling of the FPA was reintroduced because a pixel

location for each centroid must be chosen, and thus the 2 nm limit should once again

apply. The reason for the 1.6 nm resolution shown in Figure 5.7 is that with the

centroiding, 2 nm is really no longer the limit. Figure 5.8 illustrates this point. If the

centroid of λ_1 lies near the left edge of a pixel, the fifth centroid from it will lie on the

right edge of the next pixel, and thus it is possible to have two samples separated by just

1 nm. The reason that the result in Figure 5.7 doesn't achieve 1 nm resolution is that
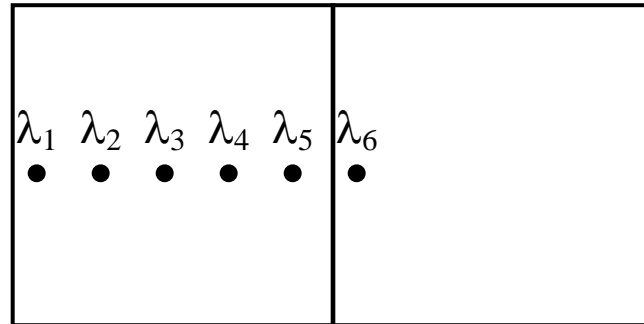
there is noise in the positions of the centroids.

Figure 5.8: Illustration that with centroiding it is possible to have two spectral samples separated by just 1 nm. In this case, the difference between λ_n and λ_{n+1} is 0.2 nm.

The limiting factor for the spectral resolution is now the FPA sampling, so

theoretically the spectral resolution of the result with simulated point sources can be

improved by upsampling the data images. Remember, this is a simulation exercise in

which both the PSF and finite spectral bandpass of the sources have been compensated

for, but is still useful for confirming the expected behavior after reconstruction. For

example, if the images are upsampled by a factor of two, there should be two samples for

every 0.5 nm and in the best case scenario quasi-monochromatic features separated by

0.5 nm could be resolved. In this case, since the centroids are nominally spaced every 0.2

nm, the experiment will only be able to show a resolution of 0.6 nm when the images are

upsampled by a factor of two. This upsampling was done, and the spectral plot of the

result for the 645-646.2 nm dual component object point source is shown in Figure 5.9.

The spectral resolution by the Rayleigh criterion is now 1.2 nm. Actually, Figure 5.9

looks a little over resolved by the Rayleigh criterion and the 1.2 nm number should be lower, however, as Appendix B shows, the next reconstruction at 1.0 nm does not show two distinct spectral components. However, the FWHM metric points to a resolution
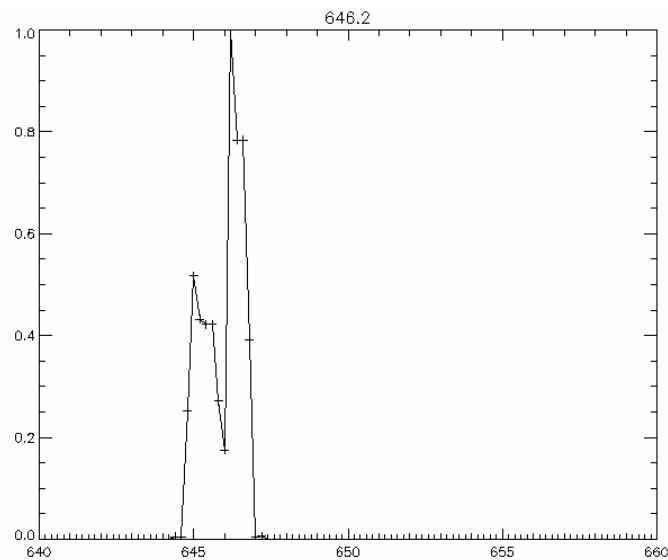

Figure 5.9: Spectral plot of the reconstruction of the 645-646.2 nm simulated point image for an upsample factor of 2.

close to 0.8 nm. The reason the Rayleigh criterion once again fails to reach the theoretical 0.6 nm (three points separated by 0.2 nm to get to the next 'pixel') is attributed to noise in the centroids' positions. The 0.2 nm translation of the spectral sources with the monochromator makes it possible to do an upsample by a factor of 4 (an expected 0.25 nm of dispersion per 'pixel') before the source spectral separation become the limit, however, the arrays become prohibitively large for reconstruction. It is also expected that the noise in the positions of the centroids has now become the real limit to

upsampling. Note that the upsampling is no longer valid when the PSF of the real CTIS system is taken into account.

Spectral resolution limit

The real goal is to simulate quasi-monochromatic point sources using centroiding in order to experimentally determine the spectral resolution limit of the instrument. In order for centroiding to be a valid representation of quasi-monochromatic point source images of a *real* system, they must be convolved with the PSF of the CTIS , shown in Figure 5.10.
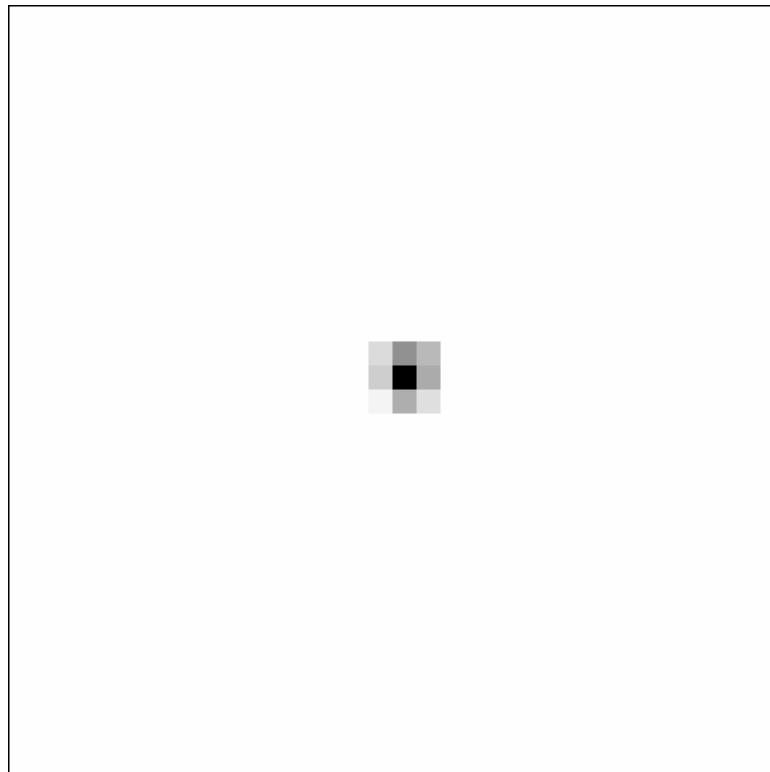

Figure 5.10: The PSF of the CTIS.

The experiment is once again repeated after convolving the simulated quasi-monochromatic point source calibration and object images with the PSF, and the result is shown below in Figure 5.11. The result confirms the geometric calculation in Figure 5.1 that the CTIS spectral resolution limit for this instrument is very near 2 nm.
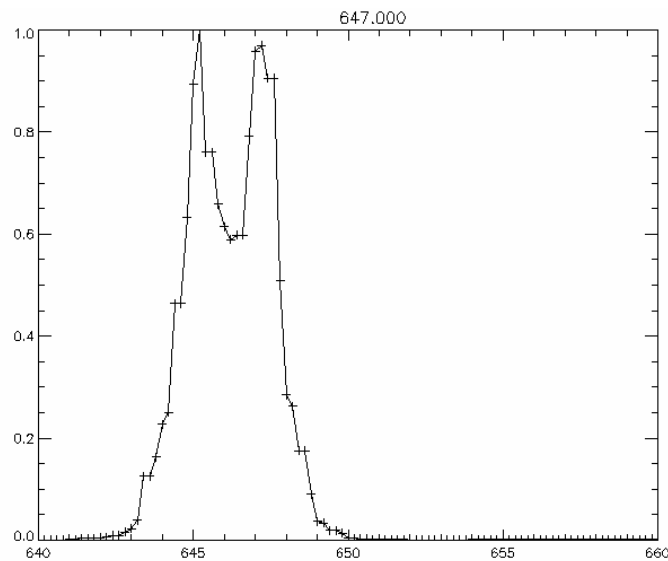

Figure 5.11: Reconstructed spectra of the 645-647 nm simulated point image with the CTIS PSF included. This represents the true experimental result of the CTIS spectral resolution limit for a 5 x 5 configuration.

This is the experimental verification of the geometric dispersion calculation, and the conclusion is that the spectral resolution limit of the CTIS is twice the dispersion calculation for the highest orders. The next question is if the spectral resolution of the tunable CTIS changes when the instrument's configuration is altered by designing the tunable disperser to change the number of diffraction orders.

Configuration comparison

The experiment was repeated using two other disperser configurations. Including the 5 x 5 configuration that has thus far been used, the three configurations are shown in Figure 5.12 using a monochromatic extended source. The first configuration is the 5 x 5 pattern, the second is a 3 x 5 pattern that omits the columns of second orders, and the third is another 3 x 5 pattern that omits the first orders. The patterns in Figure 5.12(a)-(c) will be referred to as patterns A, B, and C, respectively. The reason that the orders in the vertical direction were not changed is that the horizontal electrode LC cell of the OPA is not used due to its voltage dependent scatter. The horizontal electrodes were held to a constant voltage that minimizes the scatter and one of the PMMA etched CGHs was used for diffraction in the vertical direction. Thus the disperser was tuned only in the horizontal direction. The reason for the faint second orders on the right side of pattern B is that the electrode voltage solution that was found for this pattern happens to couple some of the light to those orders.
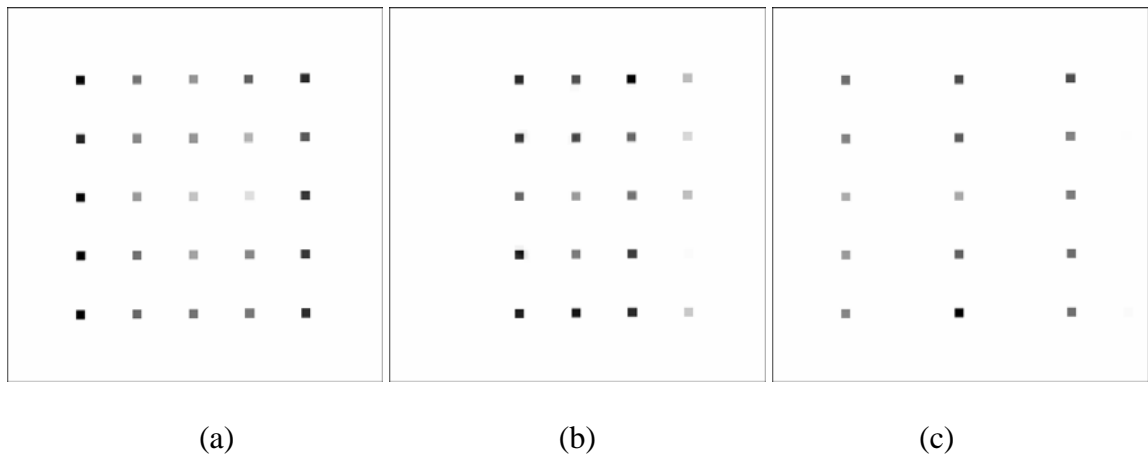

(a)   (b)   (c)

Figure 5.12: The three configurations under test: (a) pattern A, (b) pattern B, and (c) pattern C.

The results of the limiting spectral resolution, reconstruction time, and limiting

SNR for the three configurations are presented in tabular form in Figure 5.13. The

limiting SNR is defined as the SNR of the faintest diffracted order in the configuration in

question. Also, the results of the reconstructions without processing ('as calibrated') and

after centroiding and convolving with the CTIS PSF ('resolution limit') are presented.

		Spectral Resolution (nm)	Reconstruciton time (sec/iter)	Limiting SNR
Pattern A	As calibrated	6.0 nm	60.0 sec	8.69
	Resolution limit	2.0 nm	13.2 sec	
Pattern B	As calibrated	6.6 nm	57.5 sec	23.37
	Resolution limit	3.0 nm	7.9 sec	
Pattern C	As calibrated	6.0 nm	63.4 sec	10.78
	Resolution limit	2.4 nm	8.1 sec	

Figure 5.13: Results of the spectral resolution experiment for the three disperser configurations.

The reconstruction time for the experiment depends on the number of points

saved in the sparse matrix formation, and this number is radically reduced by centroiding.

Again, it is stressed that the only reason for the centroiding is that sufficient point sources

cannot be achieved with the current equipment. The implication is that the same gains in

spectral resolution and reconstruction time would be achieved without centroiding if

sufficient point sources were available.

The limiting SNR is found by taking the maximum value of the faintest order in

the 640 nm calibration image divided by the mean of the image. This maximum value

was less than one tenth the dynamic range of the camera and is an arbitrary measure, but

is consistent between the three configurations and shows a gain of about a factor of three

in SNR between the faintest projections by switching from pattern A to pattern B.

Pattern C does not show a significant SNR gain even though it has fewer projections than pattern A because the disperser was not properly optimized. In the case of pattern C, the algorithm simply didn't find the best CGH solution for this goal diffraction pattern and too much of the light was coupled into the zero diffraction order. Repeating the feedback design is time consuming, and since pattern C is sufficient for the proof of the concept its optimization was not pursued.

At first glance the results shown in Figure 5.13 are unexpected. There is actually a 50% reduction in the experimental value for the spectral resolution limit when comparing the values for both patterns A and B (from 2 nm to 3 nm). Since pattern B merely cuts out the second orders in the horizontal direction, it was expected that there would *not* be a change because there are still second orders in the vertical direction and the dispersion across the FPA per pixel is the same in all the second orders. The reason for the change is that there are simply too few second orders in pattern B - six of them as opposed to ten of them for pattern C, and sixteen of them for pattern A. There are an insufficient number of second order projections to sample the content of the object cube using pattern B – i.e. there are too few planes in the frequency representation of the data cube for pattern B (Figure 2.2a illustrates the use of a finite number of planes). Accordingly, there is a 20% reduction in the spectral resolution limit of pattern C.

The hypothesis at the beginning of this chapter was that the spectral resolution limit of the CTIS as calculated by geometric dispersion could be experimentally verified, and this was proven correct. Reconstruction affects the spectral resolution of the data cube due to the tomographic nature of the measurement, and care must be taken to

consider the CTIS configuration when deciding whether the spectral resolution can be inferred from the dispersion calculation. In the absence of adequate point sources to experimentally determine the spectral resolution limit of a CTIS, this chapter presents a method based on centroiding to simulate ideal quasi-monochromatic point sources. Only three simple conditions need to be met for this method to be a valid: (1) the point source must have a circularly symmetric irradiance distribution, (2) the CTIS PSF must be circularly symmetric, and (3) the spectral content of the point source must be symmetric about its center wavelength.

Alternative calibration method

In order to avoid simulation of adequate calibration images, the calibration sources need to be made point-like and have a very small bandpass. The major difficulty is that the signal is too low when acquiring calibration images, and this is exasperated when the bandpass of the source is further reduced. The suggestion, then, is to use lasers as quasi-monochromatic sources. There would then no longer be a need for a monochromator or to image the exit face of a fiber to the field stop. The front objective lens of the CTIS would not have to be removed, and the laser source could shine through a pinhole placed on axis and far enough away from the CTIS such that its image is the PSF of the instrument at that particular wavelength. The problem with this suggestion is that several tunable lasers[27] would be needed to accomplish the calibration and the cost would be prohibitive. However, Hege[28] has shown that only a few calibration wavelengths spanning the band of interest are needed along with Fourier interpolation in

order to construct the CTIS system **H** matrix. Laser from the green to red portion of the

visible spectrum are commercially available at low cost, and the only major cost factor

would then be in the blue. The assumption is that the calibration images are a smoothly

varying function of wavelength, which is true of the CTIS.

CHAPTER 6 – CONCLUSIONS AND FUTURE WORK

The goal of this dissertation has been twofold: (1) to integrate a liquid crystal spatial light modulator into a CTIS instrument and characterize its performance as a tunable CTIS disperser, and (2) to implement tunability by analyzing different CTIS configurations.

To realize these goals, the Raytheon Optical Phased Array (OPA) was integrated into a visible CTIS. An analysis of the OPA characterized its phase vs. voltage response and the effects of its crosstalk on a binary phase profile. The binary phase response of the OPA was also useful to characterize the transfer function of the OPA. In this way the actual phase profile of the binary phase was modeled by its spatial frequency components.

The crosstalk of the OPA forced the use of a feedback CGH design, rather than an a priori design in the computer. This lead to an analysis of the effects of nonlinearity in the current polychromatic linear inversion technique that is used to design CTIS dispersers. Initially the algorithm would not converge when implemented with the OPA in feedback. The analysis and experimentation showed that three modifications to the algorithm were necessary when relating OPA *voltages* to far-field amplitudes in the FPA plane rather than relating CGH *phases* to complex electric fields. The three modifications were to threshold the smallest singular values, damp the voltage correction vector, and strictly enforce the voltage range after correction.

Once these modifications were added, the feedback algorithm converged and the OPA was successfully designed. Using this feedback design, the OPA can be designed to perform as any CGH that could be fabricated from the current, state-of-the-art design method and then etching that design into an appropriate material.

An experiment was also performed which analyzed the implementation of tunability. It was chosen to determine the effects of altering the number of diffraction orders on CTIS reconstructions of a point source. From the analysis, a simple method for post-processing the CTIS calibration and object source images was developed to experimentally determine the spectral resolution of the instrument. It was determined that although higher orders, such as the second orders, contain the same dispersion per pixel, the resolution after reconstruction changes depending on the number of orders (i.e. projections) which exist for the reconstruction. The conclusion is that not only do the highest orders need to have sufficient spectral sampling, there must be several of them in order to have enough projections to properly sample the data cube. If there are enough of these projections, the spectral resolution limit of the final data cube can be inferred from a geometric calculation of the dispersion in the highest orders.

Future work

What remains is to utilize tunability to perform optical processing. This includes designing the OPA to whiten backgrounds so that known spectra or sets of spectra can be identified without reconstruction, modulating the efficiencies of the diffracted orders to explore the possibility of spreading information into the region of the missing cones, or

even programming a quadratic phase to the OPA concurrent with the CGH phase profile to replace the reimaging lens of the CTIS. The linear inversion technique needs to be analyzed in more depth to determine its possibilities. As Volin[9] states, "A statistical model must be generated which expresses the set of possible spectral variation which may be generated…." The suggestion here is to develop a simulation of an idealized CTIS and perform an exhaustive study of the effects of the disperser designs that can be achieved in the computer.

The advent of the next generation of LC SLMs leads to dramatic improvement of the tunable version of the CTIS. Two-dimensional, fully addressable devices with the appropriate feature sizes for the visible through the LWIR are coming to the market at reasonable costs. It is difficult to get more than a wave of phase delay out of these devices (with the current LC materials), however, they could be cascaded to allow for the dynamic phase range needed to perform optical processing with the CTIS.

The issue of CTIS calibration holds opportunity for future creativity. An exciting possibility is the use of Fourier interpolation to vastly simplify the calibration procedure in the lab, as well as open the door to the possibility of using lasers as quasi-monochromatic calibration sources

Lastly, but certainly not least, is to add computer control of zoom lenses along with a tunable disperser to achieve a fully tunable CTIS and study the possibilities of trading off spatial and spectral resolutions.

APPENDIX A – OPA COMMUNICATIONS

The Optical Phased Arrayed fabricated by Raytheon Network Centric Systems was originally designed as a portable demonstration device in the early 1990s. The electronics control box (see Figure 4.1) was custom built and converts 10 bit numbers that are passed to it from a computer via 10 base T raw Ethernet packets into actual voltages to be applied to the transparent electrodes. The computer for the portable demonstration OPA is an HP Omnibook running Windows 3.1.1 and the control software is written in LabVIEW. It is unclear what driver software actually sends the messages to the OPA electronics box, and the protocol of those messages (Ethane) is unknown. The Omnibook does not have the computing power to do the feedback SVD CGH design, so new control software was developed. A 37 pin extension cable was purchased to connect the OPA to the electronics box so that the OPA could be placed in the CTIS.

The first step was to decode the messages that are sent to the OPA via the HP Omnibook because there is no documentation available. The LabVIEW interface allows the user to select any 10 bit value for 16 different electrodes, and a message is sent every time a value is changed. A free downloadable packet sniffer, Ethereal, was used to detect the packets of each message and their contents. Each message contains five packets. The first packet is a broadcast packet, and the second packet is the response of the electronics box to the broadcaster, giving the broadcaster its source address. The next two packets contain the electrode voltage data from the computer to the electronics box, and the last packet is confirmation of receipt from the electronics box to the computer.
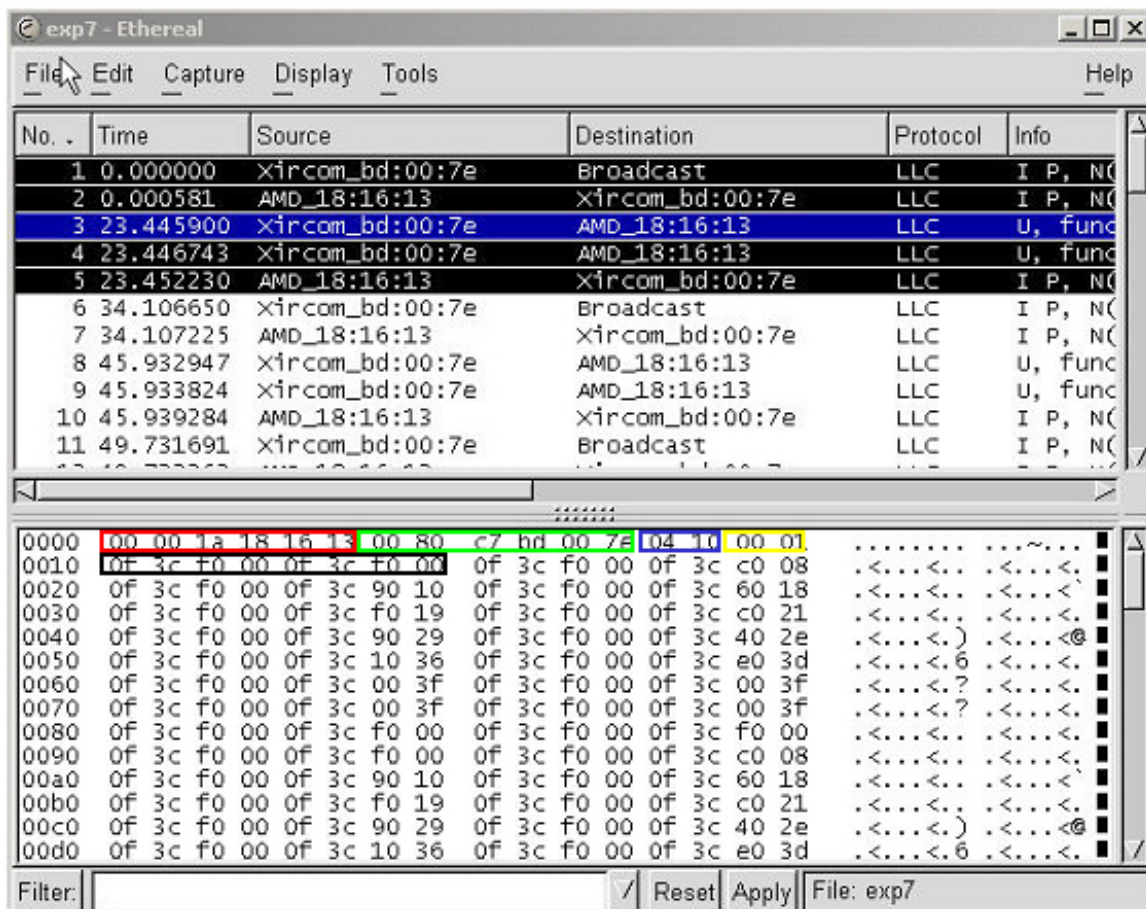
Figure A.1: Screen shot of the Ethereal packet sniffer highlighting the five packets of a message that is sent to the OPA, with the data shown for the packet highlighted in blue.

Figure A.1 shows a screen shot of the Ethereal packet sniffer output after it has detected several packets. The first five highlighted packets are the content of a single message. Of interest for the decoding are the third (highlighted in blue) and fourth packets – these contain the data for the electrode voltage numbers. The first thing to notice is that the data is split into two packets, each 1040 bytes in length, with the first 16 bytes containing the destination address (6 bytes in the red border), source address (6 bytes in the green

border), the size of the message in bytes (2 bytes in the blue border) and a non-standard packet number designation that simply is either "1" or "2" (2 bytes in the yellow border), presumably denoting the first or second data packet. All numbers are in hexadecimal format, where one byte corresponds to 2 numbers that are paired together, each ranging from 0 to f (8 bits for each byte). The next thing to notice is that although the HP Omnibook only controls sixteen of the electrodes, the message sends a voltage number to all 256 of the independently addressable electrodes – the sixteen electrodes are simply repeated. The first eight bytes of the second line, which is the start of the actual data (line 0010 in Figure A.1) contains the data for the first electrode *in 6 different channels*. This is because of the coarse/fine architecture that Raytheon uses. The first data packet contains the data for the first 128 electrodes (8 bytes per electrode – 1024 bytes), and the second message contains the data for the last 128 electrodes.

The last two channels, 5 and 6, control the coarse OPA used for this project. It is presumed that two of the first three channels control the fine OPA. It is also presumed that two of the channels are left empty, most likely for convenience in constructing the custom boards in the electronics box. All six channels can be modified in the LabVIEW software to view the effects in the packets, and it was determined that each channel contains 10 bits such that data is sent to 60 of the possible 64 bits for each electrode, and four of the bits remain unchanged. The first four bytes contain the data for channels 1-3, and the second four bytes contain the data for channels 4-6. These are independent of each other, however, the data within each of these four bytes is not in order. This is best illustrated in Figure A.2 below. The top line is the data from the sniffer – the 8 bytes

inside the black border in Figure A.1. The data is only shown for the last four bytes

(rightmost four bytes in the black border) since this is the data for channels 4-6.

0				f				3				c				f				0				0				0				Data
0	0	0	0	1	1	1	1	0	0	1	1	1	1	0	0	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	Bit values
31	30	29	28	27	26	25	24	23	22	21	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1	0	Bit number
Ch 6 - Red					5	4	3	2	1	0	15	14	13	12		---------->				0	0	0	0	0	0	1	1	1	1		15	Ch 6
Ch 5 - Green					11	10	9	8	23	22	21	20	19	18		---------->				0	0	0	0	0	0	1	1	1	1		15	Ch 5
Ch 4 - Blue					17	16	31	30	29	28	27	26	25	24		---------->				0	0	0	0	0	0	1	1	1	1		15	Ch 4
Bit number 15 actually corresponds to real number zero!! There is an offset of 15 added to the sent data																																

Figure A.2: Spreadsheet illustrating the arrangement of the data for each channel within each number that

is sent to a particular electrode.

The bits are ordered from most significant to least significant, from left to right, as

denoted by the bit number row. The colors in this row denote which channel that

particular bit belongs to, and the placement of that bit within the number for the particular

channel is shown in the middle on the left. For example, bits 0,1,2,3,4,5,12,13,14,15 all

belong to channels 6, however, they are ordered from left to right, most significant to

least, as 5,4,3,2,1,0,15,14,13,12. There is an offset of 15 for all channels. This, then, is

the electrode number "0" for all three of these channels. Channel 5 corresponds to the

horizontal electrodes as the OPA sits atop the electronics box, and channel 6 corresponds

to the vertical electrodes.

There are just two more complicating factors. First, it is expected that the

electrode number, since there is an offset, can only range from 0 to 1007. However, the

experimentation shows that the range is only from 0 to 993. There seems to be another

offset of 15 at the high end of the number as well. Second, the arrangement of the

electrode numbers is flipped in the second message. For example, the first message will

send the data to electrodes 1,2,3,4,…,128, corresponding to bytes 16 to 1040. The

second message will send the data to electrodes 256,255,254,…,129, corresponding to

bytes 16 to 1040. This is illustrated in Figure A.3, where the second data message is now

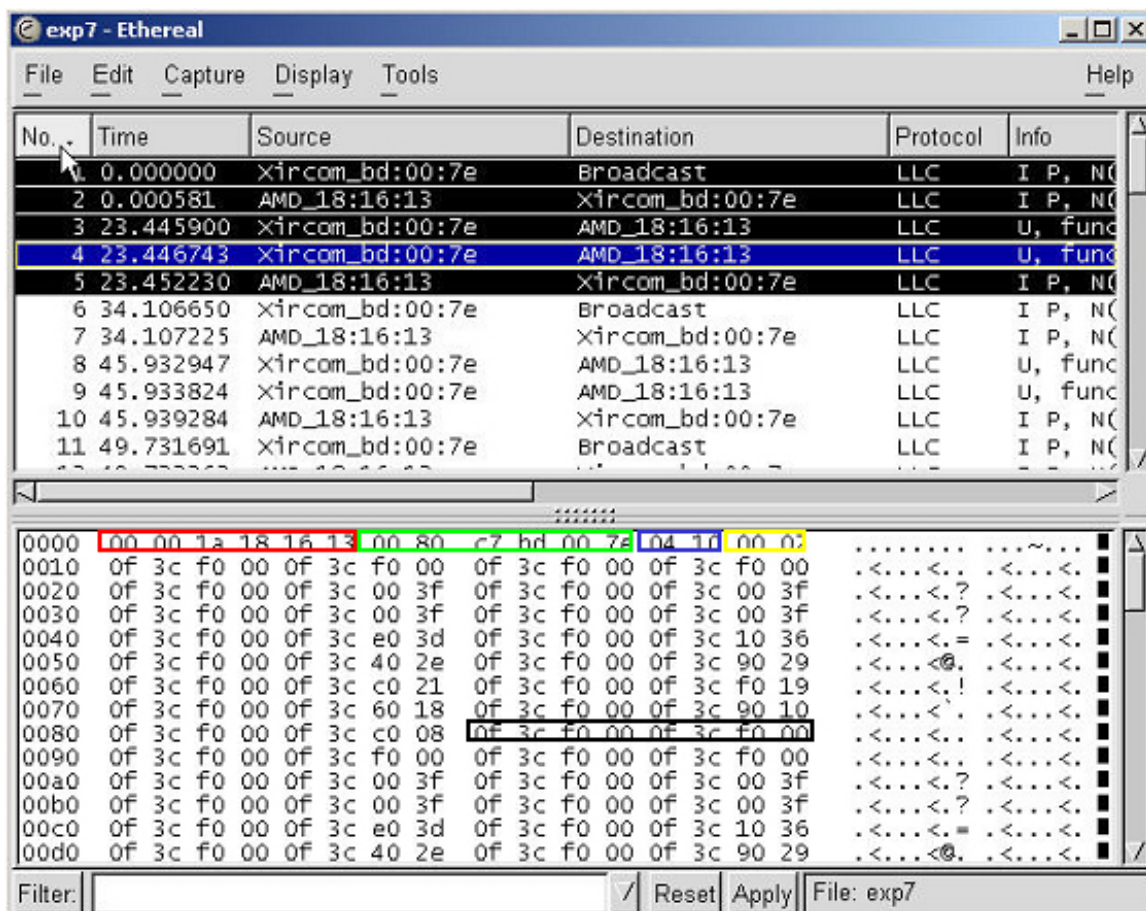highlighted. The original data values programmed to the 16 electrodes on the HP

Figure A.3: The second packet of the message illustrating the flipping of the order of the electrodes that

are addressed in the message.

Omnibook were 0, 125, 250, 275, 400, 525, 650, 725, 850, 975, 1000, 1000, 1024, 1024, 0, 0. Note that the values that actually get sent will be 0, 125, 250, 275, 400, 525, 650, 725, 850, 975, 993, 993, 993, 993, 0, 0.

After decoding, a program must be written to send the correct packets. This proved to be much more difficult than it seems. Dr. Kevin McNeill was instrumental in developing the ability to send raw Ethernet packets using contemporary operating systems (Windows NT, 2000, or XP). The first trial was to use Windows sockets, however, sockets demand some type of protocol and do not allow the user access to the Media Access Control layer (MAC).[29] Windows also provides software called Data Link Control (DLC) that allows the user access to the MAC layer in order to send raw packets, according to Windows documentation. This was concluded to be false.[30] Apparently DLC seeks to establish a connection between the sender and receiver before customized packets can be sent. The third method, which was successful, was to utilize the Network Driver Interface Specification (NDIS).[31] This requires writing drivers in Windows DDK, the driver development kit. However, these drivers do exist and one was bought from PCAUSA, a company near Atlanta, over the internet. Within two hours of purchase the software was delivered and communications with the OPA were established.

The author is truly grateful for the use of the OPA. Although the OPA was limited by its crossed 1D architecture and low transmission in the visible for use in the CTIS, it is truly an amazing device with repeatable performance and gave instant CGH gratification.

APPENDIX B – FWHM METRIC

One of the most widely used metrics for determining the resolution of a spectrometer is the full width at half maximum (FWHM) of detected spectral components. The assumption with this metric is that the components have a width that is symmetric. For the most part, the spectral components of the tunable CTIS spectral resolution experiment are symmetric and the FWHM metric can be used. However, when two spectral components are very close, the FWHM metric becomes ambiguous and instead the Rayleigh criterion was used. The following presents the results of reconstructions near the just-resolved cases presented in Figures 5.7, 5.9 and 5.11 and shows the data from which the spectral resolution was chosen as well as the FWHM of the components.

Centroided case

The reconstructions shown below in Figure B.1(a)-(d) are the companion plots to Figure 5.7, each with a slightly different spectral separation between the two components. Figure B.1(a) has a 2 nm separation, (b) has a 1.8 nm separation, (c) is a carbon copy of Figure 5.7 and has a separation of 1.6 nm, and (d) has a separation of 1 nm. The spectral resolution of lies somewhere between Figure B.1(b) and (c), and thus (c) was chosen as the closest representative of the resolution. Figure B.1(d) shows that 1 nm is not

resolvable, and Figure B.1(a) shows two components that are clearly resolved and from which a FWHM of about 1 nm can be read.
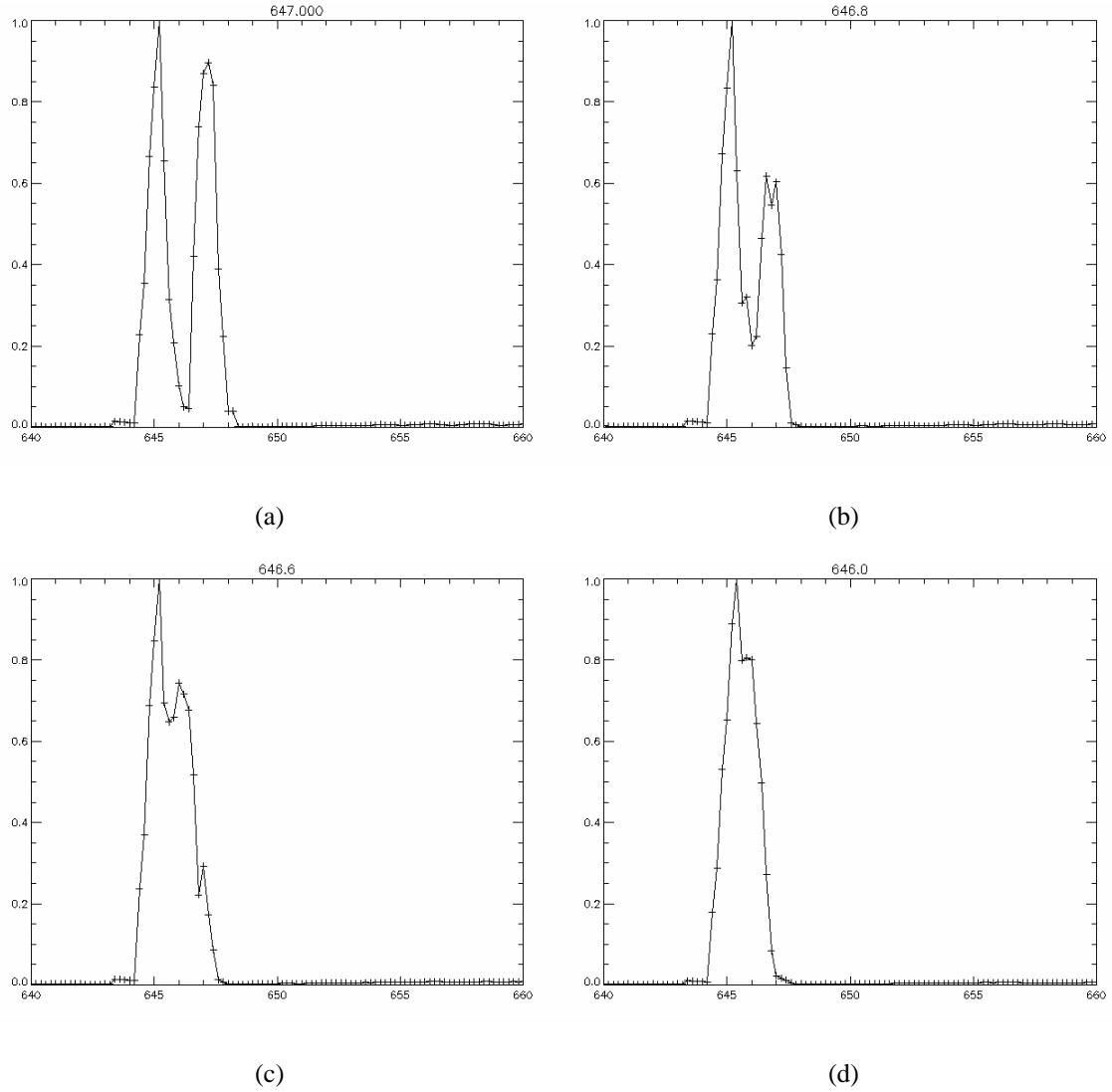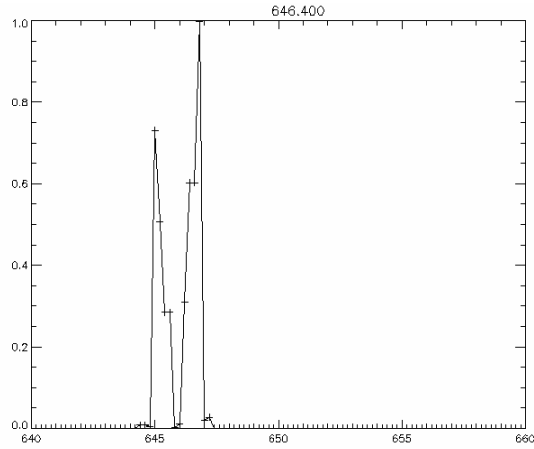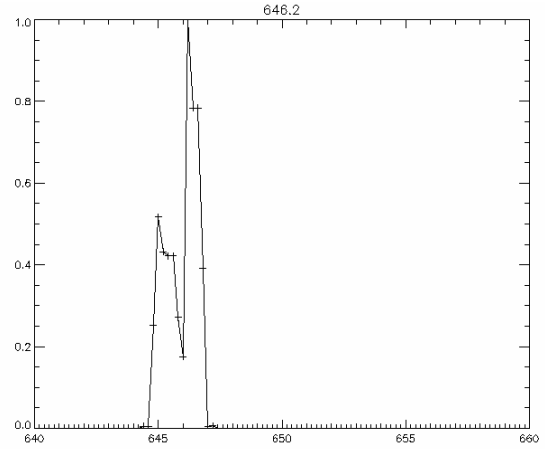

(a)   (b)

(c)   (d)

Figure B.1: Plots of the dual spectral components separated by (a) 2 nm, (b) 1.8 nm, (c) 1.6 nm, and (d) 1 nm. Plot (c) is a repeat of Figure 5.7 and is the spectral resolution of the tunable CTIS according to the Rayleigh criterion after centroiding and before convolving by the PSF.
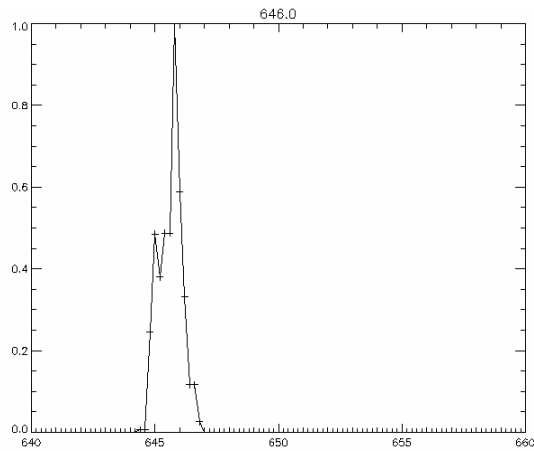
Upsampled case

The reconstructions shown in Figure B.2(a)-(c) are the companion plots to Figure

5.9, again each with a slightly different spectral separation between the two components.

(a)   (b)

(c)

Figure B.2: Plots of the dual spectral components separated by (a) 1.4 nm, (b) 1.2 nm, and (c) 1 nm. Plot

(b) is a repeat of Figure 5.9 and is the spectral resolution of the tunable CTIS according to the Rayleigh

criterion after centroiding and upsampling but before convolving by the PSF.

Figure B.2(a) has a 1.4 nm separation, (b) is a carbon copy of Figure 5.9 and has a separation of 1.2 nm, and (c) has a separation of 1 nm. In this case, Figure B.2(b) was chosen as the best representative of the Rayleigh criterion. The spectral components of the Figure B.2(a) show a FWHM of about 0.6 nm.

PSF case

The reconstructions shown in Figure B.3(a)-(d) are the companion plots to Figure 5.11, once again with each having a slightly different spectral separation between the two components. Figure B.3(a) has a 3 nm separation, (b) has a 2 nm separation, (c) is a carbon copy of Figure 5.11 and has a separation of 1.8 nm, and (d) has a separation of 1.6 nm. In this case, Figure B.3(c) was chosen as the best representative of the Rayleigh criterion. The spectral components of the Figure B.3(a) show a FWHM of about 1.4 nm.
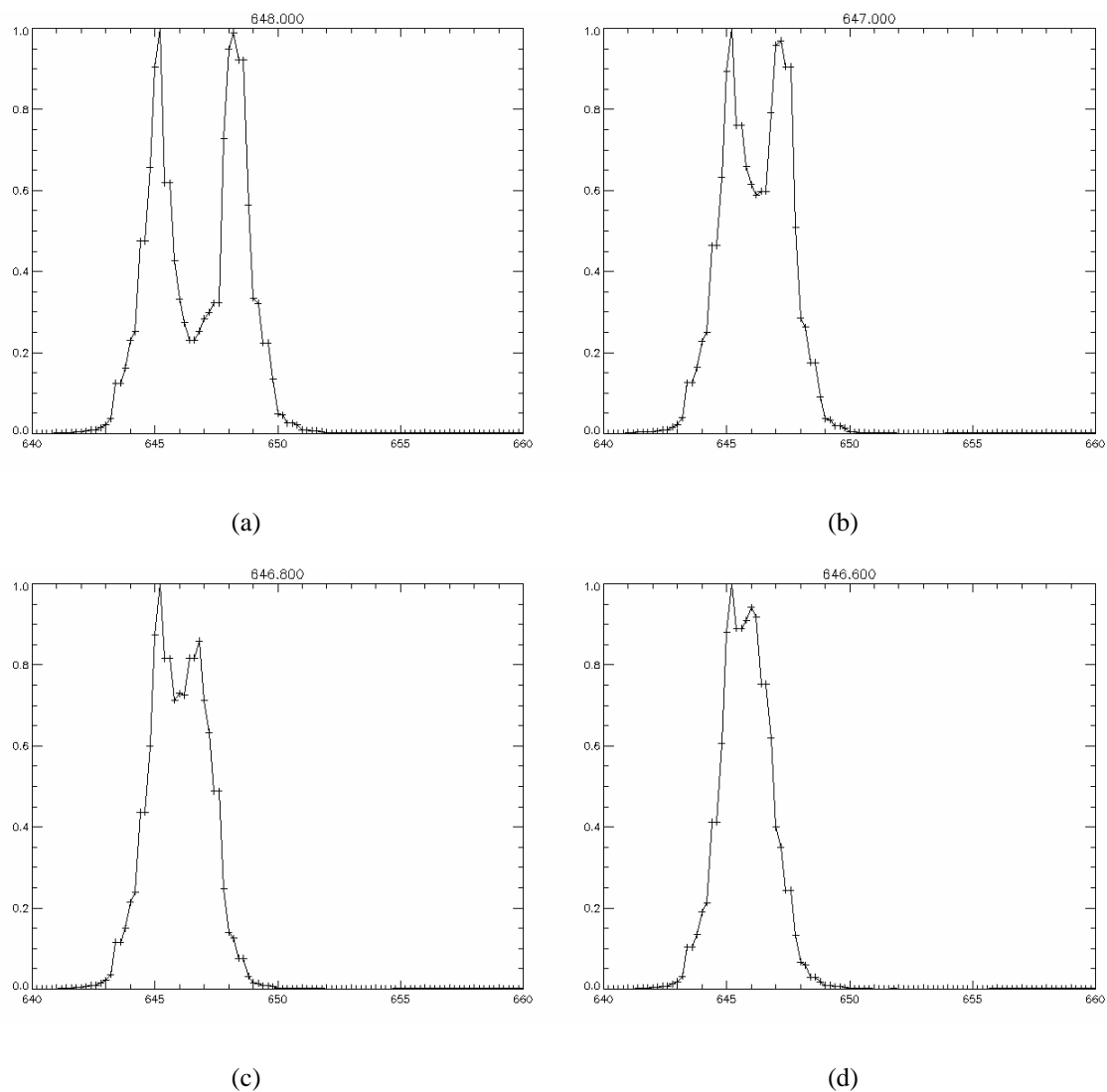
(a)

(b)

(c)

(d)

Figure B.3: Plots of the dual spectral components separated by (a) 3 nm, (b) 2 nm, (c) 1.8 nm, and (d) 1.6 nm. Plot (c) is a repeat of Figure 5.11 and is the spectral resolution of the tunable CTIS according to the Rayleigh criterion.

[1] FOVEON website: http://www.foveon.com/X3_tech.html

[2] AVIRIS website: http://aviris.jpl.nasa.gov/

[3] P. R. Norton, "Status of infrared detectors," in *Infrared Detectors and Focal Plane Arrays V*, E. L. Dereniak and R. E. Sampson, eds., Proc. SPIE **3379**, pp. 102-114 (1998).

[4] Brian Smith, *Fundamentals of Fourier Transform Infrared Spectroscopy,* CRC Press, Boca Raton, 1996.

[5] J. M. Mooney, V. E. Vickers, An Myoung, A. K. Brodzik, "A High-throughput hyperspectral infrared camera," *Journal of the Optical Society of America A*, **14**, No. 11, pp. 2951-2961 (Nov. 1997).

[6] M. R. Descour and E. L. Dereniak, "Computed-tomography imaging spectrometer: experimental calibration and reconstruction results," *Applied Optics* **34**, 4817-4826 (1995).

[7] P. F. McManamon, T. A. Dorschner, D. L. Corkum, L. J. Friedman, D. S. Hobbs, M. Holz, S. Liberman, N. Q. Nguyen, D. P. Resler, R. C. Sharp, E. A. Watson, "Optical Phased Array Technology," *Proc. IEEE* **84**, pp. 268-298, 1996.

[8] M. R. Descour, "Non-scanning imaging spectrometry," Ph.D. Dissertation, University of Arizona (1994).

[9] C. E. Volin, "Portable snapshot infrared imaging spectrometer," Ph.D. Dissertation, University of Arizona (2000).

[10] J. W. Goodman, *Introduction to Fourier Optics*, McGraw-Hill, San Francisco, Second Edition, 1996, pp. 81-83.

[11] M. S. Kim, M. R. Feldman, C. C. Guest, "Optimum encoding of binary phase-only filters with a simulated annealing algorithm," Optics Letters, v. 14, p. 545 (1989).

[12] C. L. Coleman, "Computer generated holograms for free-space optical interconnects," Ph. D. Dissertation, University of Arizona, p. 72 (1998).

[13] William H. Press, Saul Teukolsky, William Vetterling, and Brian Flannery, *Numerical Recipes in C*, pp. 408-412 (Cambridge University Press, 1992).

[14] R. W. Gerchberg and W. O. Saxton, "A practical algorithm for the determination of phase from image and diffraction plane pictures," *Optik* **35**, pp. 237-246 (1972).

[15] M. R. Descour, C. E. Volin, T. M. Gleeson, E. L. Dereniak, M. F. Hopkins, D. W. Wilson, and P. D. Maker, "Demonstration of a computed-tomography imaging spectrometer using a computer-generated hologram disperser," *Applied* Optics, **36**, 3694-3698 (1997).

[16] C. E. Volin, M. R. Descour, and E. L. Dereniak, "Design of broadband-optimized computer-generated hologram dispersers for the computed-tomography imaging spectrometer," in *Imaging Spectrometry VII*, M. R. Descour and S. S. Shen, eds., Proc. SPIE **4480**, pp. 377-387 (2002).

[17] C. L. Coleman, "Computer generated holograms for free-space optical interconnects," Ph. D. Dissertation, University of Arizona, Ch. 3 (1998).

[18] H. H. Barrett and K. J. Myers, *Foundations of Image Science*, Wiley Interscience, Hoboken, 2004, pp. 34-42.

[19] Liquid Crystal tutorial of Case Western Reserve University: http://plc.cwru.edu/tutorial/enhanced/files/textbook.htm

[20] L. M. Blinov, *Electro-optical and magneto-optical properties of liquid crystals.* Chinchester, UK: Wiley, p. 292, 1983.

[21] J. Gaskill, *Linear Systems …*

[22] J. W. Goodman, *Introduction to Fourier Optics*, McGraw-Hill, San Francisco, Second Edition, 1996, p. 102.

[23] Dr. Terry Dorschner, personal communication, July 2003

[24] Dr. Tomasz Tkaczyk, personal communication, April 2004

[25] M. Born, E. Wolf, *Principals of Optics*, Cambridge University Press, Cambridge, Seventh Edition, 1999, p. 371

[26] Lord Rayleigh, *Phil. Mag.* (5), **8** (1879), p. 261.

[27] http://www.opticsjournal.com/tutorial.htm

[28] Dr. E. Keith Hege, personal communication, January 2005

[29] Dr. Kevin McNeill, personal communication, October 2002

[30] Dr. Kevin McNeill, personal communication, December 2002

[31] Dr. Kevin McNeill, personal communication, January 2003